National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

The Structure of the Complex of Polypeptide Chymotrypsin Inhibitor 1 and *Streptomyces griseus* Proteinase B

by

Harry Mark Greenblatt

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF Master of Science

Department of Biochemistry

EDMONTON, ALBERTA

Spring 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-42891-0

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Harry Mark Greenblatt

TITLE OF THESIS The Structure of the Complex of Polypeptide Chymotrypsin Inhibitor 1 and *Streptomyces griseus* Proteinase B

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Science

YEAR THIS DEGREE GRANTED Spring 1988

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) .....

PERMANENT ADDRESS:

14 Sherwood Dr.

Ottawa Ont.

K1Y-3V5

DATED April 15 1988

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled The Structure of the Complex of Polypeptide Chymotrypsin Inhibitor 1 and *Streptomyces griseus* Proteinase B submitted by Harry Mark Greenblatt in partial fulfilment of the requirements for the degree of Master of Science.

Michael James
.....
Supervisor

Brian D. Sykes
.....

Cyril M. Kay
.....

Monica Palcic
.....

Date April 15 88

Abstract

A low molecular weight protein inhibitor of serine proteinases from Russet Burbank potato tubers, polypeptide chymotrypsin inhibitor-1 (PCI-1), has been crystallized in complex with *S. griseus* proteinase B (SGPB). The three dimensional structure of the complex has been solved at 2.1Å resolution by the molecular replacement method and has been refined to a final R-factor ($=\Sigma||F_O|-|F_C||/\Sigma|F_O|$) of 0.142 (8.0 to 2.1Å resolution data). The reactive site bond of PCI-1 (Leu38I to Asn39I) is intact in the complex, and there is no significant distortion of the peptide from planarity. The distance between the active site serine O^{γ} of SGPB and the carbonyl carbon of the scissile bond of PCI-1 is 2.8Å. The inhibitor has little secondary structure, having a three-stranded antiparallel β-sheet on the side opposite the reactive site and 4 β-turns. PCI-1 has four disulphide bridges; these presumeably take the place of extensive secondary structure in keeping the reactive site conformationally constrained. The pairing of the cysteine residues, which had not been characterized chemically, is as follows: Cys3I to Cys40I, Cys6I to Cys24I, Cys7I to Cys36I, and Cys13I to Cys49I. The molecular structure of SGPB in the PCI-1 complex agrees closely with the structure of SGPB complexed with the third domain of the turkey ovomucoid inhibitor (OMTKY3; Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420-4433). A least squares overlap of all atoms in SGPB gives an rms difference of 0.37Å. One of the loops of SGPB (Ser35 to Gly40) differs in conformation in the two complexes by more than 2.0Å rms for the main-chain atoms. Thr39 displays the largest differences with the carbonyl carbon atom deviating by 3.6Å. This conformational alternative is a result of the differences in the molecular structures of the P'_4 residues following the reactive site bonds of the two inhibitors. This displacement avoids a close contact (1.3Å) between the carbonyl oxygen of Ser38 of SGPB and Pro42I C^{β} of PCI-1. The solvent structure of the PCI-1·SGPB complex includes 179 waters, 2 sulphate or phosphate ions, and one calcium or potassium ion, which appears to play a role in crystal formation. The molecular structure of PCI-1 determined here has allowed the proposal of a model for the structure of a two domain inhibitor from potatoes and tomatoes, Inhibitor II

(Graham, J. S., Pearce, G., Merryweather, J., Titani, K., Ericsson, L. H., & Ryan, C. A. (1985) *J. Biol. Chem. 260*, 6561-6564). The two domains of the model are related by a pseudo two-fold axis between the three-stranded β-sheet of each domain.

Acknowledgements

First and foremost I must express my thanks to the Omniscient, who has granted me the privilege of seeing a small example of the wisdom of His work, for without His help, this would not have been possible.

Secondly, I would like to thank Mike James who is endowed with a wealth of knowledge, a fraction of which I have tried to mine. I hope that I have not shortened his life too much through frustration.

During my stay here there have been numerous people who have always been willing to teach, help, and put up with redundant, basic questions. Anita Sielecki, a local world expert on refinement (among other things) has provided invaluable advice and help during the various stages of this work. Osnat Herzberg and Catherine McPhalen also assisted with map interpretations and other aspects of the work. Randy Read was quite patient with my questions and somewhat half-baked ideas. The data processing could not have been accomplished without the assistance of Masao Fujinaga. Bernard Santarsiero and Marie Fraser provided useful suggestions and advice for this thesis. Koto Hayakawa prepared the samples of SGPB, and grew the crystals used for the structure determination. Gregory Pearce (laboratory of C. A. Ryan) prepared the samples of PCI-1. Much computer related assistance came from David Bacon, Mark Israel, and Mark Botkin.

Lastly I reserve special thanks for John Moult who provided very useful advice, help, and perhaps most importantly, a British sense of humour ("-Morning captain!").

I have been funded throughout most of this work by the Alberta Heritage Foundation for Medical Research.

Table of Contents

Chapter	Page
I. Introduction	1
A. Serine Proteinases	1
B. Inhibitors of Serine Proteinases	5
C. X-ray Diffraction	10
Molecular Replacement	14
Limitations of Crystallography	15
II. Experimental	17
A. Crystallization	17
B. Data Collection and Processing	17
C. Structure Determination	18
Refinement	19
III. Results and Discussion	24
A. Quality of final model	24
B. Structure of PCI-1	27
C. Reactive site	36
D. Solvent Structure	40
E. Comparison of PCI-1 to Other Solved Inhibitors	44
F. Comparison of the SGPB structures from the PCI-1·SGPB and OMTKY3·SGPB complexes	51
G. Comparison of Solvent Structure in the Two Complexes	57
H. Relationship of PCI-1 to Inhibitor II	60
IV. Mechanism of Inhibition	72
V. Conclusion	75
Bibleography	76

List of Tables

Table II.I- Data Collection and Data Processing 18

II.II- Refinement Parameters and Results for the Last Cycle 23

III.I- Summary of Main-Chain Bond Lengths for PCI-1·SGPB Complex 26

III.II- Intramolecular H-bond Distances in PCI-1 31

III.III- Summary of Turns in PCI-1 35

III.IV- Summary of Intracomplex Contacts Between PCI-1 and SGPB 38

III.V- Geometries of Reactive Site Residues in PCI-1 39

III.VI- Ligand-Sulphate Distances 43

III.VII- Ligand-Ion Distances and Coordination Angles for Ca^{++}/K^{+} Ion 45

III.VIII- Comparison of Main-Chain Conformational Angles from Several Inhibitors 50

III.IX- Summary of Symmetry Contacts in Crystals of PCI-1·SGPB and OMKTY3·SGPB 55

III.X- Summary of Symmetry Contacts in Crystals of PCI-1·SGPB and native SGPB 58

III.XI- Sequences of Inhibitor II from Tomatoes and Potatoes 61

III.XII- Alignment of Chymotrypsin Domain (PCI-1) of Inhibitor II with the Trypsin Domain 64

III.XIII- Summary of *in-vacuo* Energy Minimization Results 66

List of Figures

Figure I.1- Argand Diagram Showing Vector Nature of F*(hkl)* 13

II.1- Plot of R-factor *vs* Resolution 21

II.2- Plot of $<_{\Delta}F>/2$ *vs* $(\sin\theta/\lambda - 1/6)$ 22

III.1- Curves for Estimated Radial Errors 24

III.2- σ_A Plot 25

III.3- Residue-Averaged B-factor Plots for PCI-1 and SGPB 27

III.4- Ramachandran Plots for PCI-1 and SGPB 28

III.5a- Schematic Representation of PCI-1 30

III.5b- Hydrogen Bonds in PCI-1 30

III.6- Electron Density in the Active Site 32

III.7- C^{α}-Plot of the PCI-1·SGPB Complex 32

III.8a- Plot of β-Sheet Region in PCI-1 33

III.8b- Schematic Representation of β-Sheet of PCI-1 33

III.9- $|F_o| - |F_c|$ Plot of Region Lys31I-Lys34I 37

III.10- Plot of Sulphate Molecules 41

III.11- Plot of Cation Ligands 44

III.12- Solvent Molecules in the Active Site Region of the Complex 46

III.13- Comparison of the Reactive Site of 4 Inhibitors to PCI-1 48

III.14- Region of Tyr34-Cys42 of SGPB in Two Complexes 52

III.15- Region of Thr168-Val177 of SGPB in Two Complexes 53

III.16- $\Delta<B>$ Plot for SGPB (PCI-1·SGPB Complex with OMTKY3·SGPB Complex) 54

III.17- $\Delta<B>$ Plot for SGPB (PCI-1·SGPB Complex with Native SGPB) 56

III.18- Model of Alternate Side-Chain for P_1 Residue 59

III.19- Hydrogen Bonding Network in Trypsin Domain of Inhibitor II 67

III.20- Model of Deletion Between Gly8 and Asn9 in Trypsin Domain of Inhibitor II 68

III.21- Proposed model for Inhibitor II 69

III.22- C^{α}-Plot of Inhibitor II 70

List of Symbols and Abbreviations

a b c	unit cell axes
B-factor	isotropic thermal vibration factor
CHT	α-chymotrypsin
CI-2	chymotrypsin inhibitor-2 from barley seeds
ConA	concanavalin A
ΔG°	change in standard free energy
e^{-}	electron
F_C	structure factor calculated from model
F_O	structure factor derived from observed intensities
$F(hkl)$	structure factor as a function of Miller indeces, h, k, l
$f_{j,\theta}$	atomic scattering factor
$\langle \Delta F \rangle$	mean $\lvert F_O - F_C \rvert$
$I(hkl)$	observed intensity indexed by h, k, l ($= \lvert F_O \rvert^2$)
i	square root of -1
MIR	multiple isomorphous replacement
OMTKY3	turkey ovomucoid inhibitor, third domain
PoTI	polypeptide trypsin inhibitor
PCI-1	polypeptide chymotrypsin inhibitor-1
PSTI	pancreatic secretory trypsin inhibitor
PTI	pancreatic trypsin inhibitor
R-factor	standard crystallographic residual
rms	root-mean-square
SGPA	*Streptomyces griseus* proteinase A
SGPB	*Streptomyces griseus* proteinase B
TPCK	tosylphenylethylchloromethyl ketone
V_m	volume per unit molecular weight

λ	wavelength of x-rays used in diffraction experiment
$\rho(x,y,z)$	electron density in real space as a function of the unit cell coordinates
σ_A	phase probability parameter
σ_F	standard deviation of a given structure factor
θ	scattering angle
$\bar{v}$	partial specific volume

I. Introduction

A. Serine Proteinases

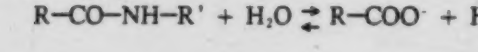
Serine proteinases are ubiquitous in nature. They function as degradative enzymes by catalyzing the hydrolysis of peptide bonds. This provides a means, for example, to convert proteins and peptides into their constituent amino acids for intake as nutrients. Proteinases are also useful for selective processing of larger polypeptide precursors to form, for instance, hormones and neurologically important peptides. The reaction catalyzed by the serine proteinases, and their postulated mechanism are described below.

On the basis of their overall three-dimensional structure, the serine proteinases can be divided into two groups, the trypsin class of enzymes and the subtilisin (formerly subtiloprotease) class of enzymes. Though these two classes, or families, have dissimilar folds of their polypeptide chains, they share several key residues which are arranged with similar spatial relationships. Three of these residues make up a catalytic triad, which is essential for the hydrolytic activity of the enzyme. In the trypsin type enzymes these are Ser195, His57, and Asp102. In the subtilisins these are numbered Ser221, His64 and Asp32. Some other common features include a site for binding the substrate with antiparallel β-sheet interactions, and two amide nitrogen atoms oriented to form hydrogen bonds with the carbonyl oxygen of the target (scissile) bond of the substrate. The former serves to orient and bind the substrate, while the latter two-fold interaction is believed to stabilize tetrahedral catalytic intermediates (Henderson, 1970), and has been named the oxyanion binding site (Robertus *et al.*, 1972).

The trypsin type enzymes are found widely in both eukaryotes and prokaryotes. Well-known examples are the enzymes, trypsin, chymotrypsin, and elastase, found in the digestive tracts of vertebrates. Bacterial examples include *Streptomyces griseus* proteinase A (SGPA) and *S. griseus* proteinase B (SGPB), which can be isolated from the extracellular fluid of cultures of this bacterium. The subtilisins, on the other hand, are found exclusively

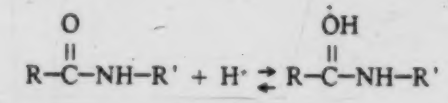
in bacilli, with one exception from fungi, proteinase K (Pähler *et al.*, 1984).

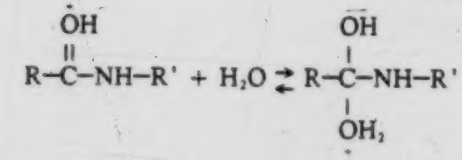
The reaction catalyzed by the serine proteinases is as follows:

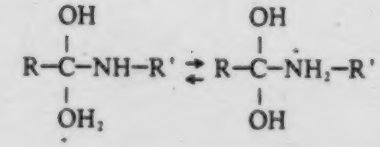


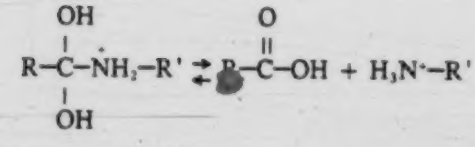
Though this is shown as an equilibrium, physiological conditions, where the products are further processed, favour the hydrolysis reaction. Furthermore, the ΔG° for the reaction is large and negative (-6.5 kcal/mole, Tinoco *et al*, 1978).

The hydrolysis reaction can occur in the absence of enzyme *in vitro*, though other catalysts are necessary to achieve appreciable approach to equilibrium. The acid catalyzed mechanism involves protonation of the carbonyl oxygen atom. This renders the carbonyl bond more electrophilic, and susceptible to attack by weaker nucleophiles such as water. The details of the reaction are as follows (adapted from Streitwieser and Heathcock, 1981):

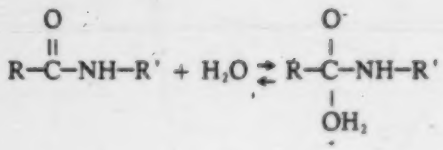




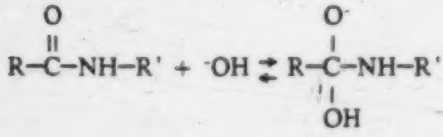


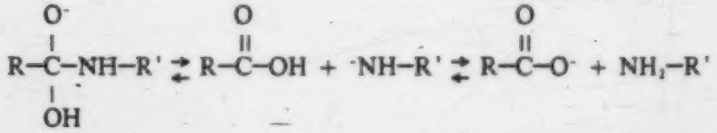


In the absence of acid, the mechanism would involve attack by a weak nucleophile, and charge separation:



Both effects are unfavourable. This is consistent with the observation that amide hydrolysis is negligible at neutral pH. In basic solution a strong nucleophile (^{-}OH) initiates the reaction:

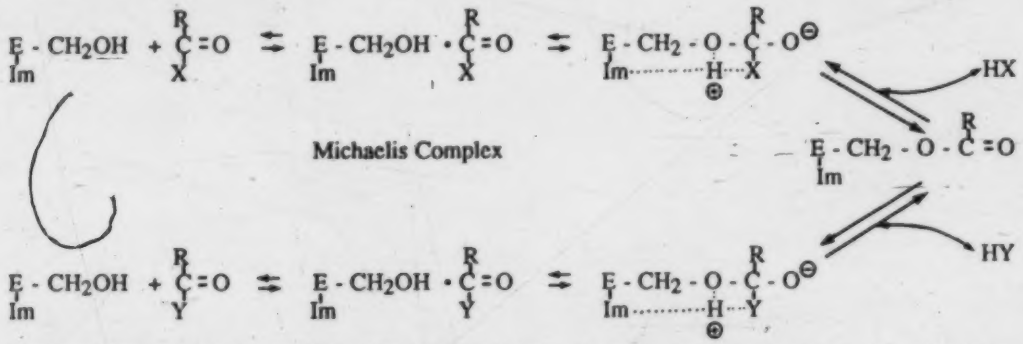




An "ideal" mechanism would presumeably involve attack by a strong nucleophile on a fairly electrophilic carbonyl group ($C=OH^{+}$). Such a combination, however, is impossible to achieve in aqueous solution, since the acidic conditions required to enhance the electrophilicity of the carbonyl group are not conducive to the formation of the hydroxide nucleophile.

The precise mechanistic details of the chemistry employed by serine proteinases to hydrolyze peptides remains unclear. An assessment of the voluminous amounts of data on this topic is beyond the scope of this thesis. Interested readers are referred to reviews by Kraut (1977) and by Steitz and Shulman (1982), as well as to papers by James *et al.*

(1980b), and by Moult *et al.* (1985). Much of the controversy revolves around the movement of electrons and the path of protons during the course of the reaction. It is generally accepted, however, that the mechanism involves formation of an ester linkage involving the substrate and the active site serine of the enzyme. The general details of the mechanism are shown here schematically (after Kraut (1977) and James *et al.* (1980b); trypsin family numbering):



Nucleophilic attack by Ser195 O^{γ} initiates the reaction; the proton from the attacking hydroxyl group is believed to be accepted by His57 $N^{\epsilon 2}$. Whether this transfer occurs before the nucleophilic attack, is concerted, or whether it occurs after covalent bond formation is a point of controversy. It is nevertheless apparent that the result is a tetrahedral enzyme adduct. The alkoxide ion is thought to be stabilized by the oxyanion binding site (nitrogen atoms of Ser195 and Gly193), and would be unprotonated.[1] Breakdown of this tetrahedral intermediate with loss of the leaving group (the P'_1 nitrogen[2] in the case of a peptide) results in the formation of an enzyme ester, commonly known as the acyl enzyme. The final step in the reaction sequence involves nucleophilic attack by a solvent molecule to give back the free enzyme and product. In a manner similar to the first half of the reaction, the

[1] Support for this hypothesis comes from the crystal structures of SGPA inhibited by peptides which are terminated by aldehyde functions rather than carboxylate groups (James *et al.*, 1980b).

[2] This is the nomenclature of Schechter and Berger (1967), which labels the substrate residues "P" and the corresponding contact sites on the enzyme "S". The residues on the acyl side of the scissile peptide are numbered P_1, P_2, P_3, etc., starting with the residue whose amide bond is cleaved. Residues towards the carboxy-terminus are labelled P'_1, P'_2, P'_3, etc., starting with the residue immediately following the cleaved peptide bond.

proton of the nucleophile is thought to be accepted by the imidazole $N^{\epsilon 2}$ of His57. If the attacking nucleophile is a water molecule, the overall reaction is hydrolysis of the amide bond of the substrate. If an alcohol is the nucleophile, the net reaction is transacylation.

It can be seen that serine proteinases appear to employ variations of the two rate enhancing conditions given above for non-enzymatic hydrolysis of peptide bonds. While these two mechanisms are mutually exclusive in aqueous solution, the surface of the protein can create isolated environments to contain the elements necessary for each type of catalysis. Thus, the peptide hydrogens of the oxyanion binding site may serve both to increase the electrophilicity of the carbonyl bond, and to stabilize the resulting alkoxide intermediate. This is similar to, but not identical with, the acid catalyzed non-enzymatic reaction where the carbonyl oxygen atom is protonated. On the other hand, the coupling of the reactive site serine, histidine, and aspartic acid side chains probably serve to increase the nucleophilicity of the serine O^{γ} through general base catalysis, creating a situation analogous to the alkaline non-enzymatic peptide hydrolysis described above.

B. Inhibitors of Serine Proteinases

Protein inhibitors of serine proteinases are a widely distributed and structurally varied group of molecules (Laskowski and Kato, 1980; Read and James, 1986). These inhibitors are fascinating in that their mechanism of action is not apparent upon inspection of their sequences, or even their tertiary structures. This is in contrast to other enzyme inhibitors which generally have more obvious modes of inhibition. Tosylphenylethylchloromethyl ketone (TPCK)[3] for example, has a very electrophilic methylene carbon (between the carbonyl group and the chlorine atom) which would be a likely candidate for nucleophilic attack. Indeed, this compound alkylates the $N^{\epsilon 2}$ of the active site histidine of serine proteinases, and irreversibly renders them inactive. The carbonyl group is also attacked by Ser195 O^{γ}, probably before alkylation of His57 occurs

[3] $C_6H_5SO_2NHCH(CH_2C_6H_5)COCH_2Cl$

(James *et al.*, 1980a).

Other inhibitors resemble either catalytic transition states, or intermediates which are structurally similar to transition states. Since the enzyme binds transition states most tightly, these inhibitors have very large association constants. The inhibition is, nevertheless, reversible with very high substrate concentrations. An example of this type of inhibitor is pepstatin (Iva-Val-Val-Sta-Ala-Sta)[4] which inhibits aspartic proteinases. The -$CHOHCH_2$- group of the statine residue resembles the tetrahedral intermediate that is formed during the course of the normal hydrolysis of a peptide bond. The inhibitors discussed in this work, however, resemble, in their primary structures, standard substrates. Even the elucidation of their tertiary structures by x-ray crystallography has not provided a clear-cut mechanistic basis for their inhibitory properties. Various hypotheses have been put forward, and these are discussed in the text.

The crystal structures of several protein inhibitors complexed to their cognate serine proteinases have been solved and refined at high resolution. These are pancreatic trypsin inhibitor (PTI, Huber *et al.*, 1974; Huber and Bode, 1978; Marquart *et al.*, 1983), pancreatic secretory trypsin inhibitor (PSTI, Bolognesi *et al.*, 1982), turkey ovomucoid third domain (OMTKY3, Fujinaga *et al.*, 1982; Read *et al.*, 1983; Fujinaga *et al.*, 1987), eglin-c from leeches (McPhalen *et al.*, 1985a, 1988; Bode *et al.*, 1986; 1987), and CI-2 from barley seeds (McPhalen *et al.*, 1985b). Many of the inhibitors can be grouped into structurally related families (Laskowski and Kato, 1980; Read and James, 1986). For example, eglin-c is similar to CI-2 in tertiary structure, and OMTKY3 belongs to the PSTI family. Despite their different tertiary structures, however, all the inhibitors have very similar conformations in the loops which bind to the respective enzymes. These loops are known as the reactive sites of the molecules. Also, all the inhibitors are wedge- or pear-shaped with the reactive site loop being the thin edge or narrow end of the molecule. Furthermore, all the inhibitors of

[4] The statine residue (Sta) is a derivative of leucine, and has the following chemical structure:

-$NHCH(CH_2CH(CH_3)_2)CHOHCH_2CO$-

known structure have a segment of the P'_n stretch of chain involved in β-sheet structure (Bolgnesi *et al.*, 1982).

Another common feature among the inhibitors is an apparently constrained reactive site. In all but two of the solved inhibitors, the reactive site is flanked by one or two disulphide bridges which are thought to limit somewhat the number of conformations available to the reactive site loop. This led to the hypothesis that the relatively rigid inhibitor does not lose as much conformational freedom, relative to a flexible substrate, upon binding to the enzyme (Blow, 1974; Read *et al.*, 1983). Thus, inhibitor binding occurs with a relatively lower entropy loss, which, coupled to electrostatic and van der Waals interactions between these highly complementary structures, gives a large and negative free energy of association. This would allow the formation of a stable inhibitor-enzyme complex. In the two cases that have no flanking disulphide bridges, eglin-c and CI-2 (McPhalen *et al.*, 1985a, 1985b; McPhalen & James, 1988), there are extensive secondary structure and electrostatic interactions which might have similar stabilizing effects. The segments of the polypeptide chain that precede and follow the reactive site loop form parallel strands in the β-sheet region of both CI-2 and eglin-c. The electrostatic interactions involve Arg65 and Arg67 which originate on the β-strand formed by the P'_n side of the reactive site loop. The guanidinium groups of these residues interact with the P_2, P_1, and P'_1 residues (McPhalen *et al.*, 1985b). Thus, in all of the inhibitors studied to date, it is believed that the reactive site loop is constrained to adopt a small group of closely related conformations, which are complementary to the active sites of possible target enzymes (Blow, 1974; Huber *et al.*, 1974; Sweet *et al.*, 1974).

The structures of some of the inhibitors have also been solved in the uncomplexed state (Weber *et al.*, 1981; Papamokos *et al.*, 1982; Bode *et al.*, 1985; McPhalen and James, 1987). In these crystal structures, the reactive site loops have large amplitudes of thermal motion relative to the rest of the molecule, as indicated by their substantially higher B-factors. The conformations of these loops, nevertheless, are essentially unchanged from

the enzyme-bound forms. This can be seen from a comparison of the bound and free forms of CI-2 (McPhalen and James, 1987). The increase in mobility possibly represents a trade off between maximally restraining the loop in a rigid conformation to permit tight binding, and allowing the loop the flexibility to adapt to similar but different proteinases (Read and James, 1986).

Polypeptide chymotrypsin inhibitor-1 (PCI-1) is one of many proteinase inhibitors of various molecular weights that have been isolated from Russet Burbank potato tubers (Hass *et al.*, 1976; Pearce *et al.*, 1982). PCI-1 is heat stable and is present in the water soluble extracts of the tubers. There are 51 amino acids in the form discussed here, and the molecular weight is 5,600. There are other isoforms which might have one or two additional residues at the C-terminus. During the initital characterization of PCI-1, only 6 cysteine residues were detected, but subsequent careful sequencing (Hass *et al.*, 1982) gave 8 cysteines in the form of 4 disulphide bridges. However, the pairing of the cysteines in the four disulphide bridges was not elucidated. The presence of 4 disulphide bridges is confirmed by the crystal structure presented here. In addition the connectivity of the cysteines has also been determined directly from the structure.

An inhibitor closely related to PCI-1, that is specific for trypsin, has also been isolated from potatoes (Pearce *et al.*, 1982). It is known as polypeptide trypsin inhibitor, PTI. This nomenclature can be easily confused with the better known pancreatic trypsin inhibitor, PTI, so in order to avoid confusion the potato trypsin inhibitor will be abbreviated here as PoTI. The sequences of PCI-1 and PoTI are 78% identical (Hass *et al.*, 1982); the P_1 leucine of PCI-1 is replaced by a lysine in PoTI. An inhibitor specific for trypsin and sharing high sequence identity with PoTI, has been isolated from eggplant (Richardson, 1979).

Another serine proteinase inhibitor from tomato leaves, called wound-induced tomato Inhibitor II, has been cloned and the cDNA has been sequenced (Graham *et al.*, 1985). The tomato inhibitor is active against both chymotrypsin and trypsin, and is 84%

indentical to an inhibitor from potatoes, potato Inhibitor II (Thornberg *et al.*, 1987). Inhibitor II is one of the major proteins in Russet Burbank potato tubers, representing about 5% of the soluble proteins. In both potato and tomato leaves, Inhibitor II accumulates in response to wounding, mediated by a systemic signal. This signal has been associated with pectic fragments from plant cell walls, apparently solubilized during pest attacks (Ryan *et al.*, 1985). The inhibitor is thought to have adverse effects on the digestive physiology of insects (Broadway *et al.*, 1986a; Broadway *et al.*, 1986b). The possible relationship between PCI-1 and Inhibitor II is presented in the discussion of this work.

Streptomyces griseus proteinase B (SGPB) is a small bacterial serine proteinase with chymotrypsin-like specificity (Jurášek *et al.*, 1974). It is isolated from the extracellular culture filtrate, Pronase. SGPB has 185 amino acids, and a molecular weight of 18,600. The structure of this enzyme has been solved by x-ray crystallography in the free state (Delbaere *et al.*, 1975) and subsequently refined at 1.7Å resolution (Sawyer, Sielecki, and James, unpublished results). The structure of SGPB has also been solved and refined in complex with the third domain of the turkey ovomucoid inhibitor, OMTKY3 (Fujinaga *et al.*, 1982; Read *et al.*, 1983). The availability of these structures contributed to the decision to solve the structure of PCI-1 in complex with SGPB. The structure of SGPB from the OMTKY3 complex proved to be an excellent search model for molecular replacement, as described below. Another contributing factor to the decision was that the PCI-1·SGPB complex would provide information about how two different inhibitors (PCI-1 and OMTKY3) interact with the same enzyme.

The sequence of SGPB has been aligned with α-chymotrypsin and its topological structure is similar to the other mammalian serine proteinases (Fujinaga *et al.*, 1985). Thus all sequence numbers used for SGPB in this paper are based upon this alignment to α-chymotrypsin; this supersedes the numbering presented in Read *et al.* (1983).

Three important conclusions have come from crystallographic studies of these enzyme-inhibitor complexes. Firstly, the scissile bond of the inhibitor is not cleaved in the

bound state, and that it is this bound form which is the stable inhibited complex. Secondly, there appears to be the possibility that some inhibitors have large scale structural flexibility to adapt to different enzyme topologies (Fujinaga *et al.*, 1987). Finally, the solvent structure in the active site of the native enzyme does not appear to be conserved upon formation of the inhibitor:enzyme complex, and the resultant solvent structure depends heavily on the structure of the inhibitor. This latter result is presented in this work.

C. X-ray Diffraction

For a detailed description of x-ray crystallography, the reader is referred to a standard text by Stout and Jensen (1968) and a work dealing with protein crystallography by Blundell and Johnson (1976).

The technique used in this work to elucidate the structure of the PCI-1·SGPB complex, x-ray crystallography, is a diffraction-based technique. Thus, the physical basis of x-ray scattering by atoms is analogous to scattering of visible light by a set of narrow slits, where the width of the slits is comparable to the wavelength of light used. If the light passing through these slits strikes a sheet of paper, one sees a series of light and dark bands. The spacing between the bands is inversely proportional to the separation between the slits. This pattern of light and dark bands is caused by constructive and destructive interference among the scattered "waves" of light as they emerge from the slits. The pattern of light and dark regions is called the optical transform of the slits, and can be derived mathematically. A suitable lens placed behind the slits would perform a reverse transformation, and give back an image of the slits. This, essentially, is the principle behind vision.

X-rays can be produced by directing a stream of electrons against a metal anode. The larger the potential of the electrons (voltage) and the greater their number (current) the more intense the generated x-rays. The wavelength, and therefore the energy, of the x-rays is characteristic of the metal used as the anode. In protein crystallography, copper anodes are used almost exclusively. The most intense emission by copper anodes is a doublet at

wavelengths 1.541Å and 1.544Å, with the former having twice the intensity of the latter. It is this doublet which is used experimentally, and the value for the effective wavelength is a weighted average of the two contributors. Since atomic interactions, such as carbon-carbon bonds (1.5Å), occur on this scale, x-rays are diffracted by atoms and molecules. Specifically, it is the electron density around the atoms which diffracts x-rays. Thus, in principle, if one could hold a molecule fixed in space and place it in an x-ray beam, one could observe the continuous transform of the molecule in three dimensions about it. As in the case of the optical transform of the slits, the x-ray transform would be made up of "bright" regions (large numbers of photons) and "dark" regions (few photons), corresponding to areas of constructive and destructive interference, respectively. If one had a lens which could focus x-rays, one could then recombine the scattered photons and produce an image of the diffracting object. Since suitable lenses for x-rays do not exist, one is forced to collect the transform.

Because molecules in solution cannot readily be held fixed, and since the diffraction caused by one molecule would be undetectably weak, crystals of proteins are used in x-ray diffraction studies for producing detailed, atomic resolution models of these macromolecules. A crystal of a protein is a regular array of molecules with the same orientation. Because of this, the transforms of individual molecules all have the same orientation and so reinforce each other. If the molecules had random orientations, as in a solution, the resultant transform would be a complex combination of the individual transforms.[5] The reinforcement of the transform also has the effect of increasing its intensity, making it detectable.

It should be noted that the three dimensional array, called a lattice, into which the molecules are packed, also has a transform. The transform of an array, in this application, is merely another array, where the distances between array nodes (lattice points) in the

[5] Solution scattering studies are possible, though they do not yield as much detailed information as high resolution crystal structures. Interested readers are referred to Cantor and Schimmel (1980).

transform are proportional to the inverses of the distances between lattice points in the original array. Because of this inverse relationship, the coordinate system of the transform is called reciprocal space, while the coordinate system of the diffracting array is called real or direct space. The effective transform that is observed in reciprocal space is a combination of the transform of the array and the transform of the molecule. Mathematically, this process of combination is referred to as a convolution of two functions. Thus, the continuous bands of varying intensity associated with the molecular transform are reduced to discrete spots by the imposition of the lattice transform. Put another way, the pattern of spots is determined by the transform of the lattice, while the intensity of each spot is determined by the intensity of the transform of the molecule at that point in reciprocal space.

It is possible to express the x-ray transform of a molecule in a lattice by the following formula:

$$F(hkl) = \sum_{j}^{N} f_{j,\theta} \exp[2\pi i(hx_j + ky_j + lz_j)]$$

Each spot in the reciprocal lattice is indexed by three integers, h, k, and l. These spots, or intensities, are referred to as "structure factors", or "F's". All N atoms in the unit cell contribute to the magnitude of each intensity. The size of the contribution from any given atom depends on the number of electrons associated with that atom, and on the angle, relative to the incident beam, at which the intensity is recorded (scattering angle). This contribution is given by the function $f_{j,\theta}$, which is a function of the number of electrons in atom j and the scattering angle, θ. This term becomes more complicated if the thermal motion of atoms is accounted for.

Note the presence of i in the exponential term of the definition of $F(hkl)$. This implies that each structure factor is a complex value. Using an Argand diagram (Figure I.1) to represent a structure factor, it is apparent that each $F(hkl)$ can be represented as a

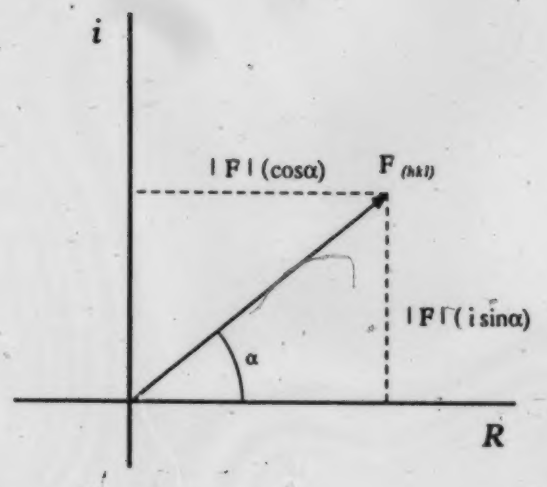


Figure I.1. Argand diagram showing F*(hkl)* as a vector quantity, with components along the real and imaginary axes.

vector, with magnitude |F*(hkl)*| and angle α relative to the real axis.[6] Using the wave analogy for light, each structure factor has amplitude F and phase angle α. The intensity of a wave is equal to the square of the amplitude, thus, I*(hkl)* = |F*(hkl)*|², or simply I = F².

The reverse transform can generate the image of the electron density from the structure factors. Thus, for any point in the crystal lattice with coordinates (*x,y,z*), the electron density is given by the expression

$$\rho(x,y,z) = \sum_{h}\sum_{k}\sum_{l} F(hkl)\, e^{-2\pi i(hx + ky + lz)}.$$

The coordinates *(x,y,z)* used in this equation are not to be confused with the coordinates used in the structure factor equation, though they both refer to real (crystal) space. The calculation of this formula requires both the amplitude and phase angle of each F*(hkl)*. In the real experiment, however, data collecting devices measure intensities (number of photons) of scattered x-rays. As stated before, the intensity is equal to the square of the

[6] The magnitude of a complex number is is given by the square-root of the product of the number with its complex conjugate. Thus, given a complex value $z = x + iy$, $|z| = zz^* = (x + iy)(x - iy)$.

amplitude, and so the square root of each intensity gives a value for the amplitude of that structure factor. The phase angle, α, for each structure factor is lost, making the reverse transform incalculable. This is known as the phase problem.

Molecular Replacement

There are various methods used to overcome the phase problem; the most common technique is known as multiple isomorphous replacement (MIR), where binding of heavy metal atoms is used to provide estimates for the phase angles. The reader is referred to Blundell and Johnson (1976) for a more detailed description of this technique.

The method employed in this work is known as molecular replacement (Rossmann, 1972). If the unknown structure, or at least a large part of it, is similar to a known structure, then the known structure can be used to supply phase information for the structure factors of the unknown molecule. The more closely related the two structures are, the more similar the phases of the model will be to those of the unknown structure, and so the better will be the correspondence between the calculated electron density and the unknown molecule. It should be noted, however, one must first ensure that the molecules are in the same orientation, both rotationally and with respect to the crystal lattice (translationally). Since movements in reciprocal space cause related movements in direct space (in the crystal), a maximal overlap of the transform of the model with the observed transform of the unknown structure should result in an overlap of the two molecules in direct space. While this is conceptually true, in practice a slightly different approach is used. Interested readers are referred to a detailed work dealing with the matter (M. Rossmann, 1972). The problem is also divided into two seperate steps: the rotational orientation is determined first, and then the translational overlap is done.

In the problem dealt with by this work, an enzyme of known structure (SGPB) was complexed with an inhibitor of unknown structure (PCI-1). Since the enzyme represented the majority of the scattering matter in the complex, it was reasonable to assume that a

properly oriented SGPB molecule would give meaningfull approximations for the phases generated by the entire complex. The model used was the structure of SGPB from the complex with OMTKY3 (Fujinaga *et al.*, 1982; Read *et al.*, 1983). The assumption was that conformational changes in SGPB caused by the binding of a different inhibitor and by changes in crystal packing would be small. This assumption proved correct, as is evidenced by the results presented below.

Limitations of Crystallography

One of the chief limitations of protein crystallography is that the collection of data occurs over a long period of time (days or weeks). As a result, motions which occur on a more rapid time frame, cause the mobile atoms to be either poorly characterized with respect to their positions, or simply invisible. Such motions generally involve side chains or entire loops on the surface of the protein. Large scale conformational changes in proteins usually present the same problem, in addition to the fact that such motion cannot usually be tolerated by the crystal packing constraints. Thus, it is often the case that crystals of native proteins which bind substrates crack or shatter when soaked in solutions of these substrates, indicating that some conformational change is occuring.

The consequence of this restriction to a static model is that many of the interesting events which occur in proteins, such as enzymatic catalysis, folding upon association of a ligand, and transport, are difficult to study by x-ray crystallography. Recent advances in millisecond x-ray diffraction, using synchrotron radiation, have allowed the study of short timescale events (Hajdu *et al.*, 1987), but this technique is still in its infancy.

Further to these considerations, protein crystallography is also limited by the fact that hydrogen atoms are not visible, owing to their weak scattering properties. These atoms do, however, contribute to the diffraction pattern, and their absence is a source of error in the models of proteins derived from such patterns. As well, the lack of hydrogen atomic positions hampers the analysis of hydrogen bonds in proteins, and so many questions remain

with regard to these important interactions. Though neutron diffraction can locate hydrogen atoms, this technique is limited by the availability of a source of neutrons (nuclear reactor).

Other problems involve obtaining the crystals themselves. The techniques for obtaining suitable crystals are highly empirical; the growing of protein crystals, then, could be quite accurately described as an art, rather than a science. For instance, limited changes to the charged group distribution on the surface of a protein may result in the requirement for drastic changes in the crystallization conditions of that protein. Furthermore, not all crystals diffract x-rays well enough to allow useful data collection.

This dissertation presents the crystal structure of the complex of polypeptide chymotrypsin inhibitor-1 (PCI-1) and *S. griseus* proteinase B (SGPB) at 2.1Å resolution. This structure is the fifth in a series of enzyme-inhibitor complexes solved in the laboratory of M. N. G. James. The aim of this series of investigations has been to elucidate the mechanism by which the inhibitor prevents itself from being cleaved at an appreciable rate. Despite the relatively large number of structures solved, no clear answer has yet emerged (Read and James, 1986). An hypothesis regarding the mechanism of action of these inhibitors is presented in the final chapter of this thesis. Site directed mutagenesis may provide important data by changing those residues thought to be vital to the function of these inhibitors.

II. Experimental

A. Crystallization

SGPB was purified as per Juràšek *et al.* (1974), and PCI-1 was purified as per Pearce *et al.* (1982). Solutions of the two molecules were mixed in a 1.7:1 molar ratio of inhibitor to enzyme. The concentration of PCI-1 was 4 mg/mL (0.71 mM) and the concentration of SGPB was 8 mg/mL (0.43 mM). The crystals of the PCI-1·SGPB complex were grown from a solution of 30% saturated $(NH_4)_2SO_4$ in 0.1 M potassium phosphate buffer and 30 μL/mL 2-methyl-2,4-pentanediol at pH 6.8-7.2 using the hanging drop technique. The crystals are rhombic in shape and have a tabular habit. The space group is $P2_1$. Unit cell dimensions are a=50.9Å, b=46.2Å, c=52.5Å, β=117.08° as measured from 18 reflections in the range $16° < 2\theta < 22°$. The volume per unit molecular weight (V_m) is 2.27Å³/Da. If a partial specific volume, υ, of 0.74 is assumed for the complex, and there are two complexes per unit cell, this corresponds to 46% of the volume of the unit cell (or 50,400Å³) being occupied by water. This amounts to approximately 850 water molecules per asymmetric unit.

B. Data Collection and Processing

X-ray intensity data were collected on an Enraf-Nonius CAD4 diffractometer. 12,941 unique reflections were corrected for absorption (North *et al.*, 1968) to a maximum value of 2.12, as well as for decay (Hendrickson, 1976), and for Lorentz and polarization. The program ORESTES (Thiessen and Levy, 1973) gave an absolute scale factor of 0.086 and mean isotropic thermal vibration factor (B) of 12.8Å². See Table I for a summary of the crystal data.

Table II.I: Data Collection and Data Processing

diffractometer	Enraf-Nonius CAD4
incident beam	Ni-filtered CuK_α, 40 kV, 26 mA
diffracted beam	60-cm crystal to counter He-filled beam path
scan type	ω scan, continuous
scan width	0.5° at 0.67°/min
background measurement	0.125° in ω on either side of peak scan
background correction	measured background corrected for intensity dependence and averaged over ranges of 2θ and ϕ
max. absorption correction (North *et al.*, 1968)	2.12
decay correction (Hendrickson, 1976)	empirical function of time and resolution, 2θ max. 10%
geometric correction	Lorentz and polarization factors

C. Structure Determination

Molecular replacement was used to solve for an initial set of phases for the complex. The search model was SGPB by itself, using the coordinates from the OMTKY3·SGPB complex (Read *et al.*, 1983). The fast rotation function (Crowther, 1972; program modified by E. Dodson) was used with the search model in a 60Å X 60Å X 60Å triclinic P1 cell to eliminate the generation of symmetry contacts. The radius of integration was from 6 to 21Å. The final orientation chosen gave a peak which was 10.7σ above the mean. The translation search algorithm was a correlation coefficient search of $|F|^2$ (BRUTE; Fujinaga and Read, 1987); the highest peak was 8.8σ above the mean. Since $P2_1$ is a polar space group, the search was done in the y=0 plane using the approximately oriented SGPB molecule as the search model. A fine search (5 dimensional) was done to determine the best orientation for the model. Coefficients used for all subsequent electron density maps were derived using a

method to reduce model bias in the maps (Read, 1986a). The MMS-X graphics system (Barry *et al.*, 1976) with the molecular modelling program by C. Broughton (Sielecki *et al.*, 1982) was used for fitting density, and UCSD MMS by S. Dempsey (running on a Silicon Graphics Iris 3030) was used for general structure analysis. The program TOM* was used for the model building described below.

The initial electron density map showed the reactive site of PCI-1 from Ala35I[1] to Asp41I quite clearly, but the density quickly became uninterpretable after these residues. The disulphide bridges from Cys36I and Cys40I were also clear, but the direction of the chain on the other sides of these cystine bonds could not be determined unambiguously. Through a combination of fitting and rephasing most of the sequence of PCI-1 was placed in electron density. A minimap contoured on stacked acetate sheets was used at one point to sort out some of the more confusing density, especially around Cys6I and Cys7I. Refinement of the model was begun when all but eight residues of PCI-1 had been fit. These eight residues were Pro1I, Lys11I, Lys31I-Lys34I, and Pro50I-Arg51I. One segment of SGPB had also been remodelled (Tyr34-Arg41), and one segment was left out (Arg48-Tyr50) at the start of the refinement.

Refinement

The restrained-parameter least-squares refinement program of Hendrickson and Konnert (1980), modified locally by M. Fujinaga for compatability with a Floating Point Systems' FP-164 attached processor, was used to refine the structure of the complex. The protocol followed in the course of the refinement is described in detail in Sielecki *et al.* (1979), and in James and Sielecki (1983). Thus, the restraints were initially relatively weak, to prevent the model from falling into a local minimum. This danger exists by virtue of the incompleteness of the model. As the refinement progressed, and the model became more

---

* This is a new version of the program FRODO (originally written by A. Jones) written specifically for the later model Iris workstations, by C. Cambillau.

[1] Throughout this work, the residues of PCI-1 are distinguished from those of SGPB by the addition of an "I" following the sequence number.

accurate and complete, the restraints were tightened to impose better geometry. For the first 60 cycles, 5.0 to 2.1Å resolution data was used. After cycle 60, the low resolution data were extended to 8.0Å.

The program to generate electron density map coefficients uses a statistical algorithm (Read, 1986a), that requires the inclusion of weak data. To maintain consistancy between the data used for calculating maps and the data used for refinement, no cutoff was used in the F_o values supplied to the refinement program. Negative F_o values were set equal to zero for both programs.

A plot of the R-factor *vs* resolution is shown in Figure II.1. The profile presented is typical. Low resolution data, with significant scattering from unmodelled bulk solvent and omitted hydrogens show poor agreement. High resolution data tend to be weak, and so have larger associated errors.

The weighting scheme used for the structure factors changed as the refinement progressed. Initially, the weighting was independant of the scattering angle, thus $\omega = 1/(\sigma_F)^2$, where $\sigma_F \simeq \langle |F_o| - |F_c| \rangle /2$. Starting with refinement cycle 61 (R-factor = 0.165), the value of σ_F was a function of resolution: $\sigma_F = C_1 + C_2[(\sin\theta)/\lambda - 1/6]$. The slope of the line was defined such that greater weight was given to higher resolution data. As the refinement progressed, the values of the constants C_1 and C_2 were adjusted so that the difference between σ_F and $\langle \Delta F \rangle /2$ was minimized for these data. This was done by plotting $\langle \Delta F \rangle /2$ *vs* $[(\sin\theta)/\lambda - 1/6]$, and fitting a straight line to the higher resolution points (3.0 - 2.1Å) with linear regression. Initial values for C_1 and C_2 were 13.8 and -28.3, respectively; final values were 12.2 and -29.7. Figure II.2 shows the plot used to derive these final values.

Solvent molecules were given variable occupancy factors after cycle 60. When these were refined, the B-factors were held fixed, and vice-versa. See Table II.I for a summary of the refinement results. The small rms shift in coordinates (~ 1/10 of the minimum estimated error), indicates that in all practicality, the refinement had converged.

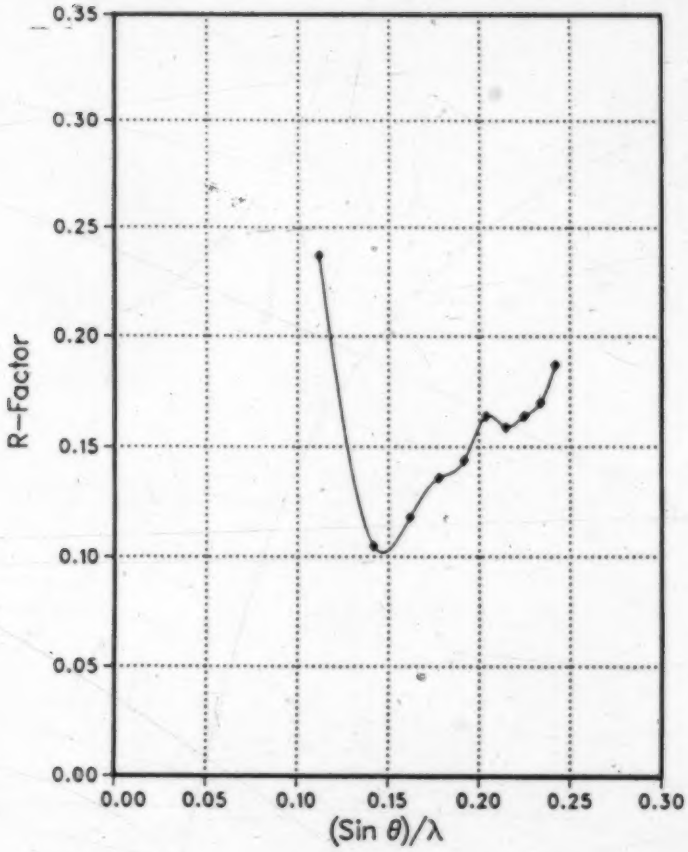


Figure II.1. Plot of R-Factor *vs* resolution for all data (no σ cutoff). Each point is the higher resolution limit for a range of $(\sin\theta)/\lambda$. Thus, the first point represents data from 0.0 to 0.112Å^{-1} (∞ to 4.5Å), etc.

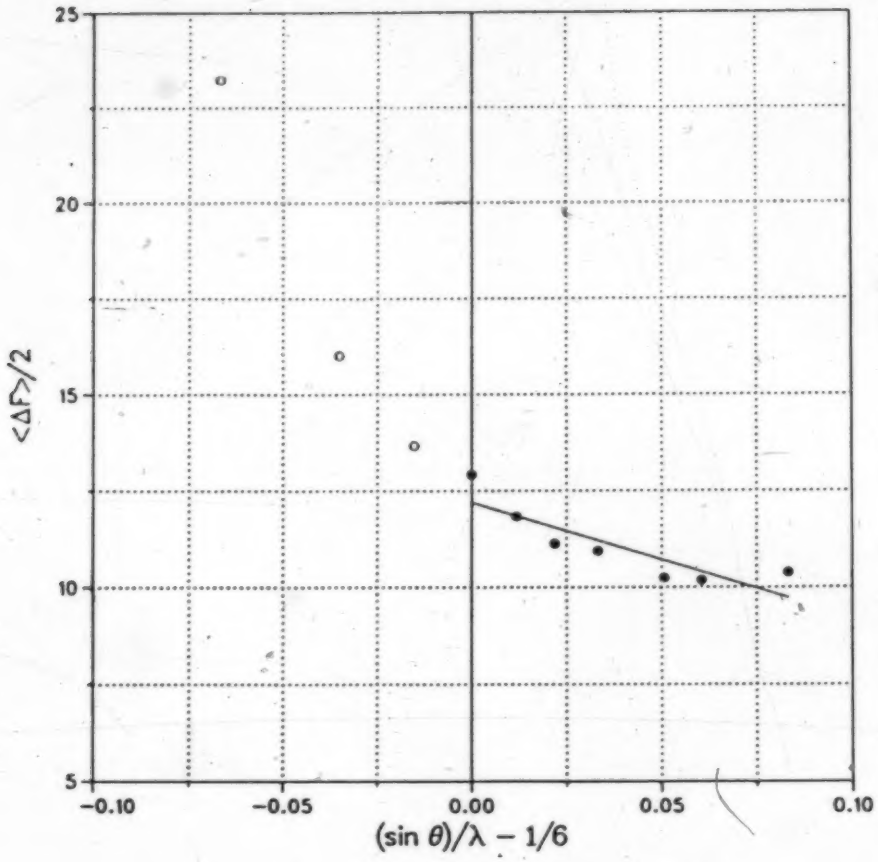


Figure II.2. Example of plot used to derive the constants C_1 and C_2 in the equation $\sigma_F = C_1 + C_2[(\sin\theta)/\lambda - 1/6]$, where σ_F is the reciprocal of the weight used for a given F_o during refinement. The line is fit to the higher resolution points (3.0 to 2.1Å, filled circles), while the lower resolution points (open circles) were not included in the regression analysis.

Table II.II: Refinement Parameters and Results for the Last Cycle

	PCI-1•SGPB	OMTKY3•SGPB[a]
no. of cycles	129	
$R(=\Sigma\|\|F_o\|-\|F_c\|\|/\Sigma\|F_o\|)$	0.142	
no. of reflections (resolution range)	12685 (8-2.1Å)	
no. of protein atoms	1690	
no. of solvent atoms[b]	190	
no. of variable parameters	7711	
coordinate shift (Å) in final cycle	0.011	
B factor shift ($Å^2$) in final cycle	0.22	
rms deviations from ideal values[c]		
distance restraints		
bond distance	0.014 (0.015) Å	0.017 (0.008) Å
angle distance	0.040 (0.030) Å	0.037 (0.016) Å
planar 1-4 distance	0.037 (0.030) Å	0.042 (0.016) Å
plane restraint	0.014 (0.015) Å	0.021 (0.012) Å
chiral-centre restraint	0.213 (0.150) $Å^3$	0.192 (0.080) $Å^3$
nonbonded contact restraints		
single torsion contact	0.203 (0.300) Å	0.259 (0.400) Å
multiple torsion contact	0.153 (0.300) Å	0.155 (0.400) Å
possible hydrogen bond	0.143 (0.300) Å	0.201 (0.400) Å
conformational torsion angle restraint		
planar (ω)	2.4 (2.5)°	3.8 (2.8)°

[a] Read *et al.*, 1983.

[b] The value shown is the sum of 179 waters, 1 Ca^{++}/K^{+} ion, and 2 sulphate and/or phosphate ions.

[c] The values of σ, in parentheses, are the input estimated standard deviations that determine the relative weights of the corresponding restraints [see Hendrickson & Konnert (1980)].

III. Results and Discussion

A. Quality of final model

The method of Cruikshank (1947, 1967) was used to provide an estimate for the radial coordinate error as a function of B-factor for different atom types in the final model of the PCI-1·SGPB complex (Figure III.1). These estimates for the radial coordinate errors of the atoms in the PCI-1·SGPB complex are comparable to those obtained for the OMTKY3·SGPB complex (Read *et al.*, 1983). The validity of the application of these formulae to protein crystallography has been discussed elsewhere (Read *et al.*, 1983). The method of Read (1986a) can be used to obtain a value for the rms coordinate error of the model. Figure III.2 shows a plot of $\ln(\sigma_A)$ *vs* $[(\sin\theta)/\lambda]^2$. The slope of the line is equal to -26.3189 * (the mean square coordinate error). The slope of the linear regression line is

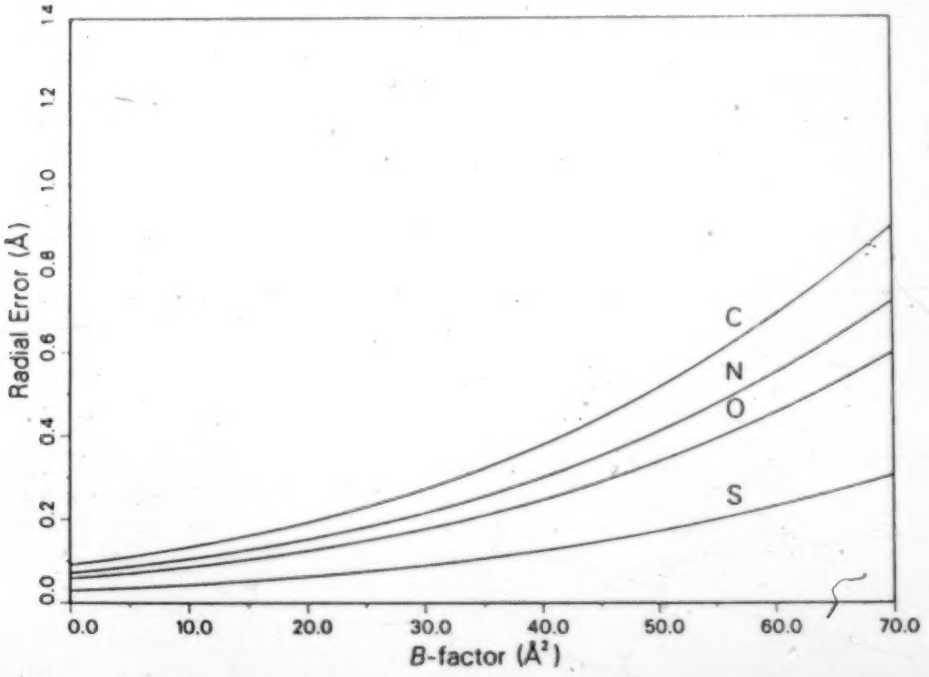


Figure III.1. Curves showing the estimated radial error in atomic positions, based on atom type, as a function of B-factor. Produced using the equations of Cruickshank (1949, 1967); C = carbon, N = nitrogen, O = oxygen, and S = sulphur.

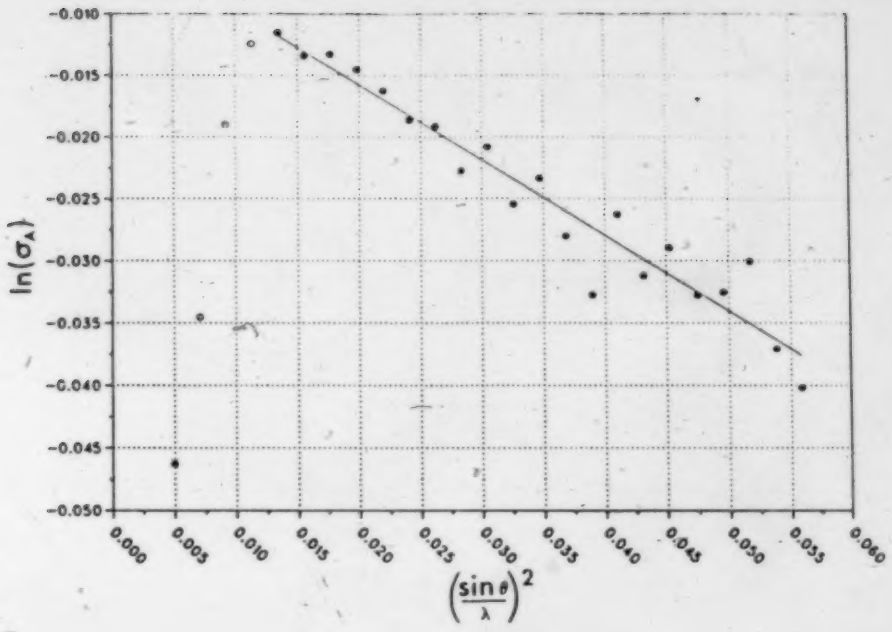


Figure III.2. Plot of $\ln(\sigma_A)$ *vs* $[(\sin\theta)/\lambda]^2$ based on the method of Read (1986). The y-intercept ($= 0.5\ln[\Sigma_P/\Sigma_N]$) has a value of 0.00216, so that $\Sigma_P/\Sigma_N = 1.004$.

-0.723 and so the rms coordinate error is 0.166Å. Dividing this value by a factor of $(3\pi/8)^{1/2}$ gives a mean coordinate error of 0.153Å (Chambers and Stroud, 1979; Read, 1986b).

In Table III.I, a summary of the main-chain bond lengths shows that there is good agreement with expected values.

The average B-factor for each residue is shown plotted against sequence number in Figure III.3. There are some differences between the profile for SGPB shown here, and that obtained for SGPB in the complex with OMTKY3 (Read *et al*, 1983). These differences are discussed in detail when the structures of SGPB in the two inhibitor complexes are compared.

Figure III.4 shows a Ramachandran-type plot for SGPB and PCI-1 seperately. The SGPB plot contains one non-glycine, non-asparagine residue which lies in the region of positive ϕ values. This is Arg80 (ϕ,ψ = 62°,52°), which has relatively high B-factors, even

Table III.I: Summary of Main Chain Bond Lengths for PCI-1·SGPB Complex

	Number	Ideal Value	Mean Value	rms[a]	Max.	Min.
N-C$^{\alpha}$	236	1.47Å	1.47	0.02	1.5	1.4
C$^{\alpha}$-C	236	1.54Å	1.53	0.02	1.6	1.5
C=O	236	1.23Å	1.25	0.01	1.3	1.2
C-N	234	1.33Å	1.32	0.01	1.4	1.3
τ[b]	236	109.7°	109.4	3.9	120.	98.3
ω (peptide)	233[c]	180.0°	-179.5	2.4	-175.2	171.3

[a]Calculated with respect to the ideal values.

[b]Angle formed by N-C$^{\alpha}$-C.

[c]Excluding *cis*-peptide before Pro95 of SGPB.

for the main chain atoms (25Å^2 for the α-carbon), and is found on an external loop of the enzyme. One of the asparagine residues, Asn100 (ϕ,ψ = 83°,-60°), occupies the third position of a reverse open turn (Ramachandran & Mitra, 1976) and immediately follows Pro95 which is part of a *cis*-peptide. A similar turn occurs in SGPA (Sielecki *et al.*, 1979), and in α-lytic protease (Fujinaga *et al.*, 1985).

Though no cutoff was used for the F_O values in the refinement, the quality of the final model is comparable to that of the final model for the OMTKY3·SGPB complex (Read *et al*, 1983), which was refined using $I \geq 2\sigma_I$, for 6.0-1.8Å resolution data. In Table II the results for the OMKTY3·SGPB refinement are presented with the PCI-1·SGPB refinement summary.

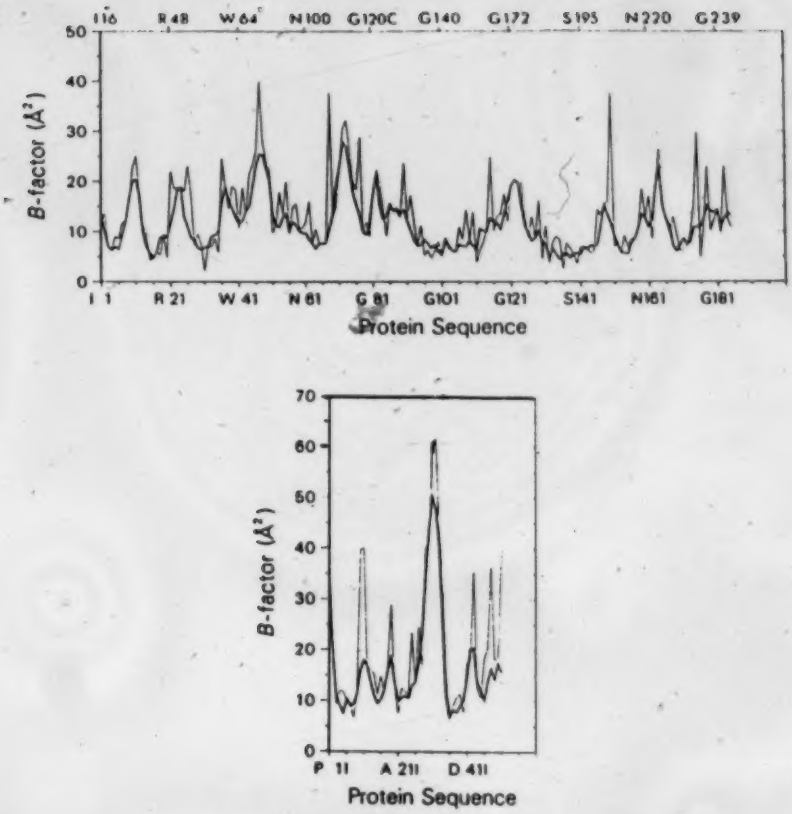
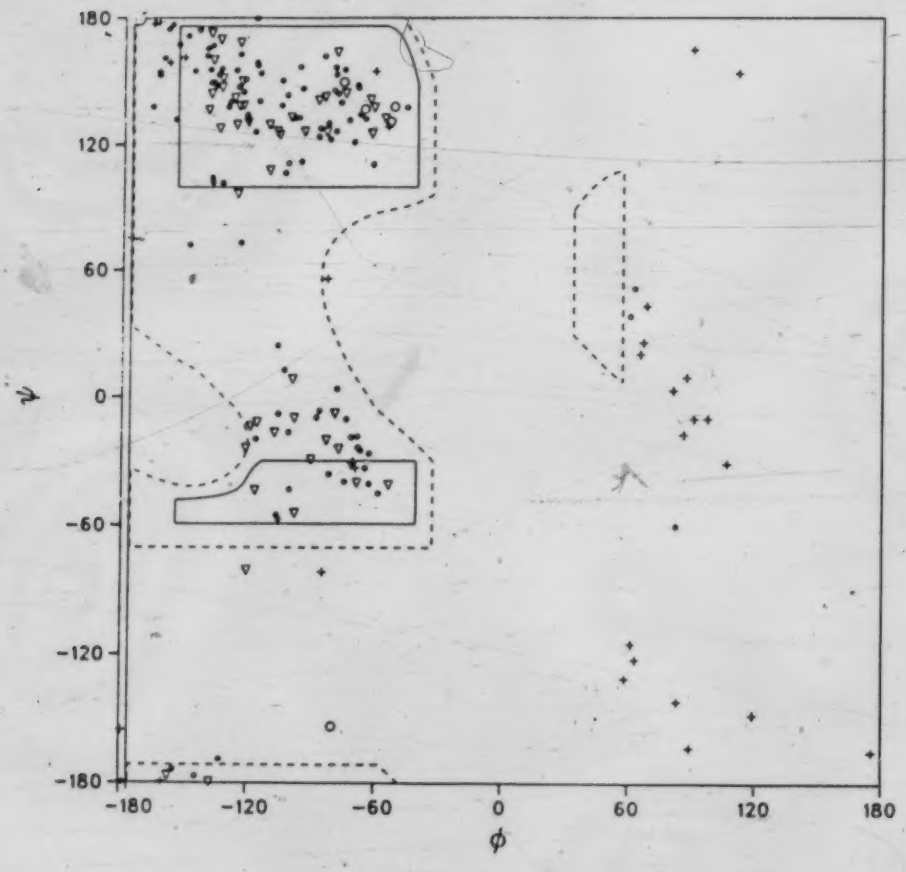


Figure III.3. Plot of average B-factor for each residue *vs* sequence number. The thick line represents the average values for the main-chain atoms, and the thin line represents the average values for the side-chain atoms. (A) SGPB molecule, with chymotrypsin numbering at top. (B) PCI-1 molecule.

B. Structure of PCI-1

PCI-1 is a thin, disk-shaped molecule, with little secondary structure. A schematic drawing of the inhibitor (Figure III.5a) shows the overall fold of the polypeptide chain. Also presented is a stereo image (Figure III.5b) showing all the atoms, and all the intramolecular hydrogen bonds[9] in PCI-1. Table III.II lists these hydrogen bonds.

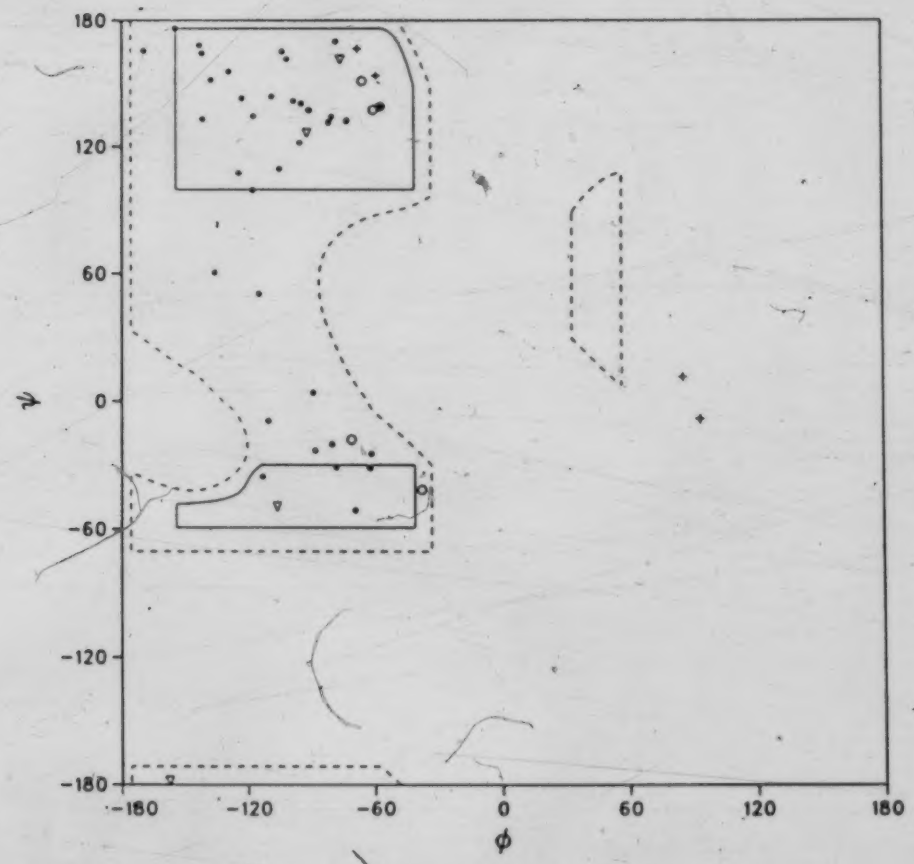
[9] The definition used for hydrogen bonds was based strictly on geometric considerations, using a program developed by M. Fujinaga. The geometric definitions used are discussed in the review by Baker and Hubbard (Baker and Hubbard, 1984). The maximum donor to acceptor distance was 3.6Å and the maximum hydrogen to acceptor distance was 2.6Å. The angle formed by the donor-hydrogen-acceptor was allowed to take on any value, but no interactions that satisfied the distance constraints had values less than 110°.



A

Figure III.4. (ϕ,ψ)-Plots for PCI-1·SGPB complex. The residues are divided into four types: glycine (+), proline (○), β-branched side chains (▽), and others (•). The continuous lines enclose the fully allowed conformational regions that are obtained by using normally accepted van der Waals' contact radii and a τ angle (N-C^{α}-C) of 115°. The dashed lines enclose the extended allowed regions when the contact distances are decreased to the absolute minimum found in small molecules (Ramakrishnan and Ramachandran, 1965). (A) SGPB. (B) PCI-1.

The extended segment of polypeptide chain from Ala35I to Pro42I is the reactive site of PCI-1; the peptide linkage between Leu38I and Asn39I is the reactive, or scissile, bond. The side chain of Leu38I fits into the hydrophobic S_1 specificity pocket of SGPB. A close non-bonded contact is made between Ser195 O^{γ} of SGPB and Leu38I C of PCI-1 (2.8Å). The electron density of this region is presented in Figure III.6, and an α-carbon atom representation of the complex is shown in Figure III.7.



B

Figure III.4 (cont'd)

The most prominent piece of secondary structure[10] present in PCI-1 is a region of antiparallel β-sheet on the side opposite the reactive site (Figure III.8). This sheet has 3 strands, with Asn14I to Tyr16I lying in the middle. This stretch of chain is flanked by Phe22I to Glu25I on one side and by Tyr46I to Lys48I on the other. Ile23I, Cys24I, and Tyr15I are involved in a "classic" type β-bulge (Richardson *et al.*, 1978), where Ile23I is residue 1 (ϕ,ψ = -106,-49), Cys24I is residue 2 (ϕ,ψ = -169,166), and Tyr15I is residue x (ϕ,ψ = -116,135). The carbonyl oxygen atom of Tyr15I forms hydrogen bonds to the amide hydrogen atoms of both Ile23I and Cys24I (Figure III.8), as is sometimes observed for this type of β-bulge.

[10] The program of Kabsch and Sander (1983) was used to help locate regions of defined secondary structure in PCI-1.

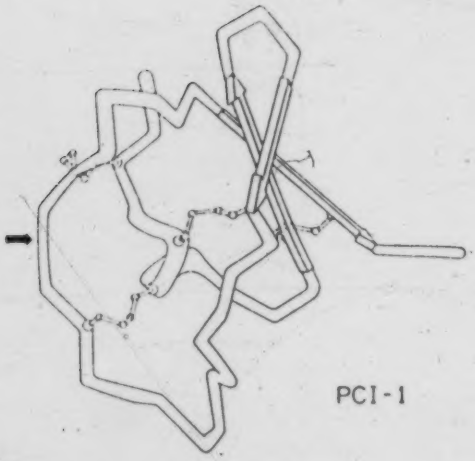


Figure III.5a. Schematic representation of PCI-1 showing secondary structural elements, disulfide bridges and the C^{α} backbone. The thick arrow shows the location of the reactive site loop.

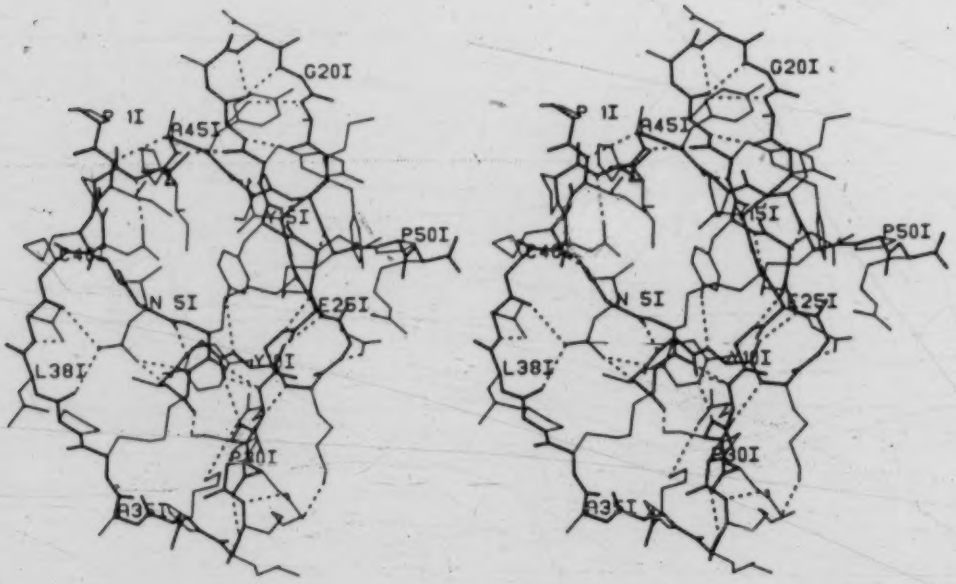


Figure III.5b. All-atom stereo pair, in an orientation similar to that shown in Figure III.5a, with main-chain bonds drawn in thick lines and side-chain bonds drawn in thin lines. Hydrogen bond interactions are represented by dashed lines. One of the hydrogen bonds is not shown for the sake of clarity. This interaction is Cys24I NH to Tyr15I O. Every fifth residue is labelled.

Table III.II: Intramolecular H-bond Distances in PCI-1 (Å)

Oxygen as Donor

Donor	Acceptor	d(O-A)	d(A-H)
S28I CB - O^{γ}	... C 7I O	2.6	1.7
S47I CB - O^{γ}	... S47I O	3.1	2.2

Nitrogen as Donor

Donor	Acceptor	d(N-A)	d(A-H)
I 2I N	... A45I O	3.5	2.5
N 5I N	... T 4I $O^{\gamma 1}$	3.0	2.4
N 5I $N^{\delta 2}$	... P37I O	2.9	2.1
N 5I $N^{\delta 2}$	... N39I O	3.0	2.0
C 7I N	... N 5I $O^{\delta 1}$	3.1	2.5
A 8I N	... N 5I $O^{\delta 1}$	2.9	1.9
G 9I N	... N 5I O	2.9	2.0
Y10I N	... Y15I O^{η}	2.9	2.0
K11I N	... S28I O	2.9	1.9
G12I N	... E25I $O^{\epsilon 2}$	2.6	1.6
C13I N	... Y10I O	2.8	2.0
N14I N	... K48I O	2.8	1.8
Y15I N	... C24I O	2.8	1.8
Y16I N	... Y46I O	2.9	1.9
S17I N	... A21I O	2.8	1.8
N19I N	... S17I O^{γ}	3.4	2.6
G20I N	... S17I O	3.0	2.0
A21I N	... S17I O^{γ}	3.1	2.2
I23I N	... Y15I O	2.7	1.7
C24I N	... Y15I O	3.4	2.4
G26I N	... C13I O	3.0	2.0
Q27I N	... C 6I O	3.0	2.2
Q27I N	... G 9I O	3.1	2.6
S28I N	... G 9I O	2.9	2.0
K31I N	... D29I $O^{\delta 1}$	3.1	2.2
K32I N	... D29I O	2.7	1.9
N39I N	... N39I $O^{\delta 1}$	2.9	2.1
D41I N	... T 4I O	2.8	1.8
H43I N	... D41I $O^{\delta 1}$	2.8	1.8
A45I N	... Y16I O	2.8	1.8
Y46I N	... Y16I O	3.1	2.2
K48I N	... N14I O	3.0	2.1

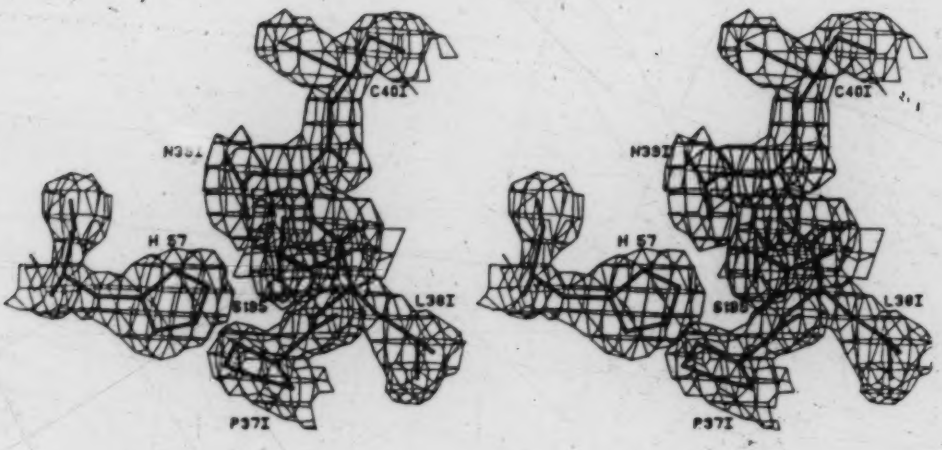


Figure III.6. $2|F_o| - |F_c|$ density of the reactive site of PCI-1, with His57 and Ser195 from the active site of SGPB included. The contouring was constrained to a 1.4Å radius around each atom. Contour level shown is $0.45e^-/Å^3$.

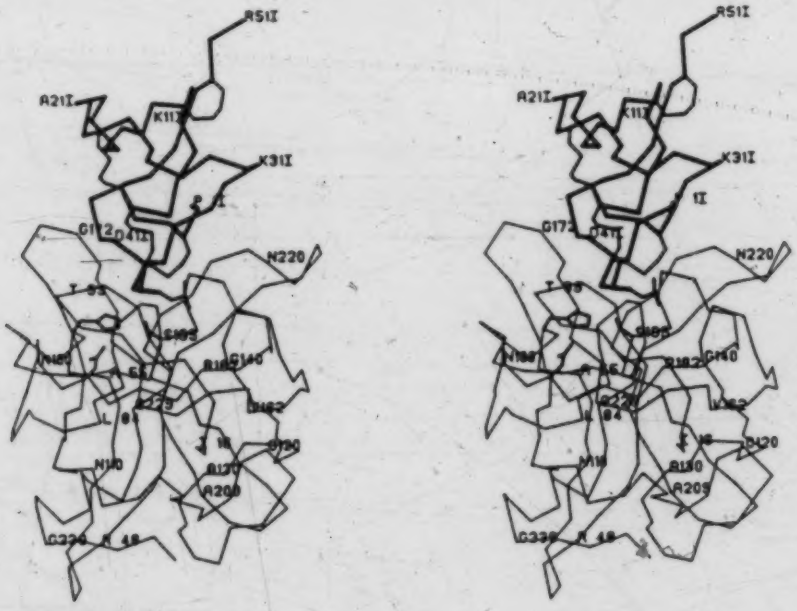
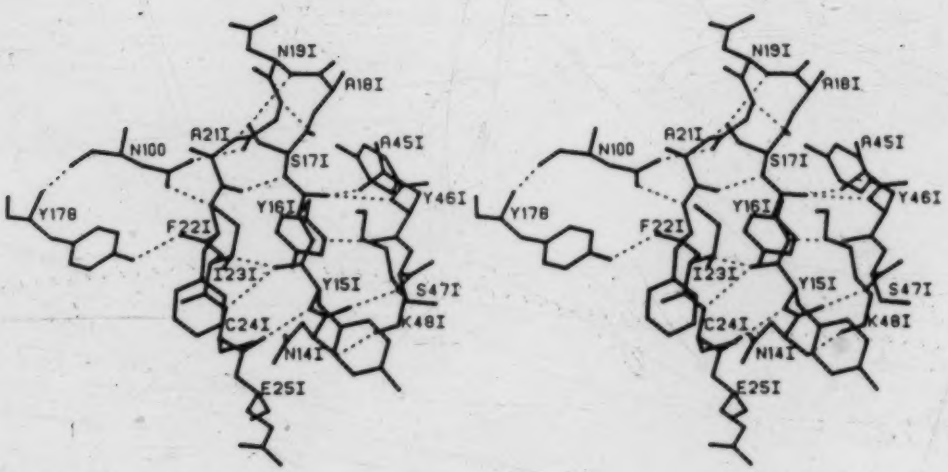
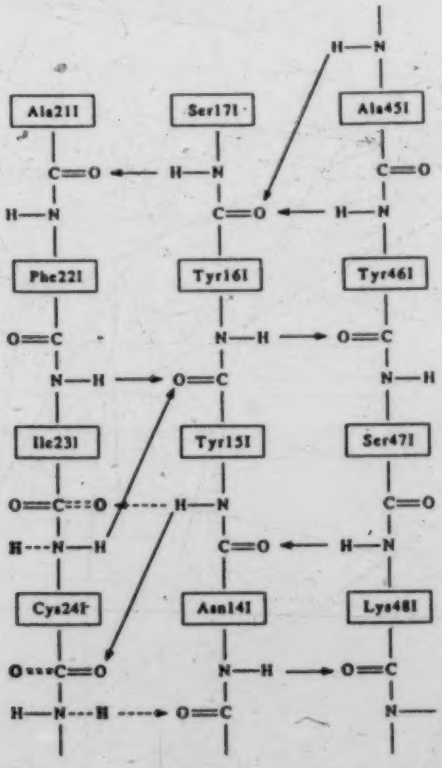


Figure III.7. C^α plot of the PCI-1·SGPB complex. PCI-1 is drawn with the thickest lines. The disulphide bridges of PCI-1, the side chain of Leu38I, and the three catalytic residues of SGPB are drawn in medium thickness. SGPB is drawn in thin lines. Every tenth residue is labelled.



A

Figure III.8. Representations of the β-sheet region of PCI-1. (A) Stereo plot showing hydrogen bonds in dashed lines. The two SGPB residues, Asn100 and Tyr178, are from a (-x,y+1/2,-z+1) symmetry neighbour. (B) Schematic drawing showing the actual hydrogen bonding pattern in solid lines and letters, and the ideal associations in dashed lines and hollow letters.



B

A type I β-turn from Ser17I to Gly20I connects two of the strands in the β-sheet (Table III.III). This turn is interesting in that the residue in postion four is the first residue of a G1 β-bulge (Richardson *et al.*, 1978). Gly20I and Ala21I are residues 1 and 2 of the bulge, respectively, while Ser17I is residue *x*. The (ϕ,ψ) values for Gly20I are (86°,11°) and for Ala21I are (-79°,135°). Also noteworthy is the fact that Ser17I O^{γ} interacts with Ala21I NH, thus providing further stabilization. An analysis by Wilmot and Thornton (1988) shows that position 1 of a type I turn is more often occupied by either aspartic acid, asparagine, serine, or threonine, and that the side chains of these residues form hydrogen bonds to the amide hydrogen of the $i+2$ residue. Those turns involving aspartic acid or asparagine are referred to as "Asx-turns" (Rees *et al.*, 1983).

Cys6I and Cys7I are in a helical conformation: ϕ and ψ for Cys6I are -62° and -31° respectively, and -69° and -51° respectively, for Cys7I. Since this 'helix' extends for just 3 residues (Cys6I to Ala8I C^{α}), other regions of the inhibitor associate with the carbonyl oxygen atoms and amide hydrogen atoms to form the favourable electrostatic interactions normally found in helices. Thus, the carbonyl oxygen of Asn5I forms a hydrogen bond with the amide hydrogen of Gly9I, the carbonyl oxygen of Cys6I interacts with the amide hydrogen of Gly27I, and the carbonyl oxygen of Cys7I forms a hydrogen bond with the O^{γ} of Ser28I. In addition, the amide hydrogen atoms of Cys7I and Ala8I both interact with the $O^{\delta 1}$ of Asn5I. It is important to note that an analogous interaction exists in OMTKY3 (Asn36I NH to Asn33I $O^{\delta 1}$; Read *et al.*, 1983). Immediately following this section of chain is a type II β-turn with the carbonyl oxygen of Tyr10I forming a hydrogen bond with the amide hydrogen of Cys13I (Figure III.5b)

There are two more reverse turns in the molecule, both of which have Asx-type interactions as well. Furthermore, both have a proline residue in position 2 of the turn. These turns with proline in position 2 have been defined as type III turns (Baker and Hubbard, 1984). The first is from Asp29I to Lys32I, which has the carbonyl oxygen atom of the former interacting with the amide hydrogen of the latter. In addition to this, the $O^{\delta 1}$ of

Table III.II: Summary of turns in PCI-1'.

Res 1	Res 2	Res 3	Res 4	d(1-4)Å	Donor	Acceptor	Type	d(Asx)Å	Donor	Acceptor
Asn5I	Cys6I	Cys7I	-	-	-	-	-	3.1	Cys7I N	Asn5I $O^{\delta 1}$
Tyr10I	Lys11I	Gly12I	Cys13I	2.8	Cys13I N	Tyr10I O	II	-	-	-
Ser17I	Ala18I	Asn19I	Gly20I	3.0	Gly20I N	Ser17I O	I	3.4	Asn19I N	Ser17I O^{γ}
Asp29I	Pro30I	Lys31I	Lys32I	2.7	Lys32I N	Asp29I O	III	3.1	Lys31I N	Asp29I $O^{\delta 1}$
Asp41I	Pro42I	His43I	Ile44I	(3.6)*	(Ile44 N)	(Asp41I O)	(III)	2.8	His43I N	Asp41I $O^{\delta 1}$

* This value is longer than the cut-off used to define hydrogen bonds; the conformation assumed is that of a Type III turn.

Asp29I forms a hydrogen-bonded Asx-turn with the amide hydrogen of Lys31I. The second type III turn occurs from Asp41I to Ile44I, with a calculated interaction energy of only -0.6 kcal/mol for the main-chain hydrogen bond (Kabsch and Sander, 1983). The side chain interaction between $O^{\delta 1}$ of Asp41I and the amide hydrogen of His43I is thus likely to have an important stabilizing effect. See Table III.III for a summary of the reverse turns in PCI-1.

The stretch of chain from Pro30I to Lys34I has very poor density associated with it; it was not possible to fit the intervening residues until after refinement cycle 21. The B-factors for the main-chain atoms in this loop are the highest in the molecule, ranging from 38Å^2 to 50Å^2 (Figure III.3) The side chains of Lys31I and Lys32I have very little density associated with them, and are probably very mobile. To test the validity of this part of the model, the residues Lys31I to Lys34I were removed at the end of the refinement, three cycles of refinement were run, and a difference map calculated. The electron density map in this region is shown in Figure III.9.

C. Reactive site

There are a total of 103 contacts less than 4.0Å between the inhibitor and the enzyme (Table III.IV), excluding those associated with crystal packing. Ninety percent of these occur between reactive site residues of PCI-1 and the active site of SGPB. One set of contacts removed from the active site occurs between Arg41 of SGPB and the disulphide bridge from Cys40I to Cys3I of PCI-1. There appear to be favourable electrostatic interactions in addition to the van der Waals forces, involved in the contact between the guanidinium group of Arg41 and the cystine bond. A similar association is seen in *S. griseus* proteinase A (SGPA; Sielecki *et al.*, 1979) between the side chain of Arg158 and the cystine group from Cys189 and Cys220 (numbering after Fujinaga *et al.*, 1985).

The distance of the active site Ser O^{γ} (Ser195) to the carbonyl carbon of Leu38I is 2.8Å, similar to values in other inhibitor-enzyme complexes (Bolognesi *et al.*, 1982;

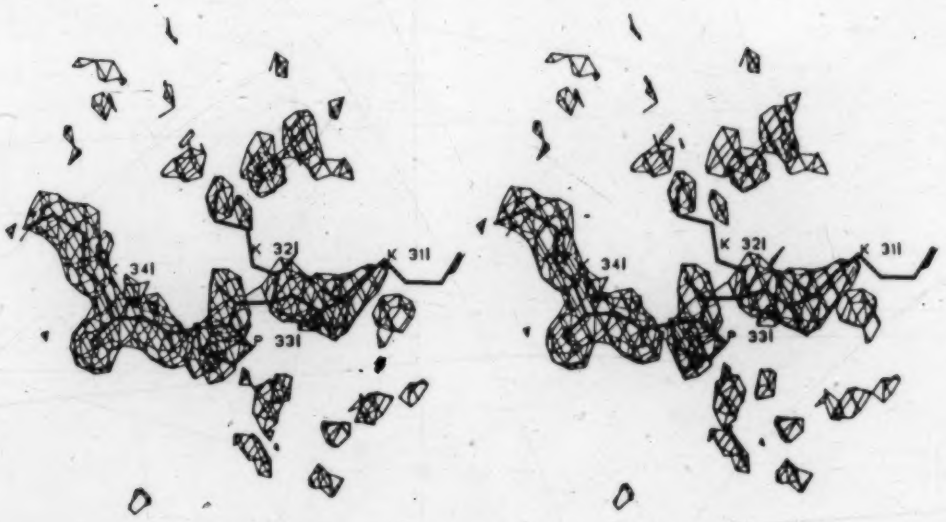


Figure III.9. Difference density ($|F_O| - |F_C|$) showing fit of residues Lys311 to Lys341 of PCI-1. The map was calculated by removing this section of chain from the final model, refining 3 cycles, and calculating difference density coefficients based on the incomplete model. The contour level for the map is $0.13e^-/Å^3$.

Marquart *et al.*, 1983; Read *et al.*, 1983; McPhalen *et al.*, 1985b; Fujinaga *et al.*, 1987). Furthermore, the planarity of the scissile peptide bond from Leu381 to Asn391 does not appear to be distorted. Evidence for this is presented in Table III.V, where the planarity of the scissile peptide bond is compared to the planarity of other peptides in PCI-1 with similar B-factors. The most distorted peptide bond in PCI-1 is from Glu251-Gly261, with a θ^4 angle of 4°.[11] This corresponds to a deviation of 0.03Å for the carbonyl-carbon atom from the plane defined by the C^α, O, and N. The largest deviation from planarity for the entire PCI-1·SGPB complex was for the peptide bond from Thr137 to Arg138 of SGPB (-6°). The possibility that restrained least-squares refinement could mask true distortion has been discussed at length elsewhere (Read *et al*, 1983; Read and James, 1986).

The above results differ from those obtained for the PTI·bovine trypsin (BT) complex (Marquart *et al.*, 1983) that was refined with the EREF program (Jack and Levitt,

[11] This angle measures the deviation of the carbonyl bond from the plane defined by the α-carbon, the carbonyl-carbon, and the amide nitrogen. A perfect tetrahedron has a θ^4 value of -54°.

Table III.IV: Summary of intracomplex contacts between PCI-1 and SGPB

	Pro 33I P_6	Lys 34I P_5	Ala 35I P_4	Cys 36I P_3	Pro 37I P_2	Leu 38I P_1	Asn 39I P'_1	Cys 40I P'_2	Pro 42I P'_4	Σ
Ser38									2	2
Thr39								1	3	4
Gly40								2	1	3
Arg41							2	10		12
Cys42							1			1
His57					2	1	3			6
Val169			1							1
Tyr171		1	3	4	5					13
Ala190						1				1
Glu191						3				3
Pro192						3	1			4
Gly193						4	1	2		7
Asp194						1				1
Ser195						10	2			12
Ser214					1	1				2
Gly215				2		2				4
Gly216		1	2	6		1				10
Ser217	3	3								6
Gly219D	1									1
Σ	4	5	6	12	8	27	10	15	6	93

An additional 10 inhibitor-enzyme contacts are made among residues Cys-3I, Thr-4I, and Ala-8I of PCI-1 to Arg-41, Pro-192, and Gly-193 of SGPB.

1978). A tetrahedrally distorted geometry for the scissile bond was reported, with a θ^* value of -11°. It should be noted that Asn233 of BT had the largest distortion for the whole complex (Read and James, 1986), with a displacement of 0.177Å. Similarly, in a recent publication of the structure of eglin-c bound to subtilisin Carlsberg, Bode *et al.* (1987) reported that the scissile peptide bond of eglin-c was deformed considerably from planar geometry. This result was achieved, however, by running three additional cycles at the end of the refinement process with all potential energy terms restraining the bond and torsion angles of the scissile peptide set to zero. Thus, as was done for PTI (Marquart *et al*, 1983), the peptide bond was afforded freedom not allowed to any other peptide group in the

Table III.V: Geometries of Reactive Site Residues in PCI-1

Residue	Bond Angles (°)				Deviations from LSQ Plane (Å)					
	$C_{\alpha i}$-C_i-O_i	$C_{\alpha i}$-C_i-N_{i+1}	O_i-C_i-N_{i+1}	C_i-N_{i+1}-$C_{\alpha i+1}$	$C_{\alpha i}$	C_i	O_i	N_{i+1}	$C_{\alpha i+1}$	rms
Cys36I	120	114	126	122	0.035	0.036	0.004	0.041	0.038	0.034
Pro37I	124	115	122	119	0.025	0.006	0.009	0.044	0.034	0.028
Leu38I	121	116	122	121	0.008	0.011	0.002	0.007	0.008	0.008
Asn39I	118	116	126	120	0.024	0.015	0.003	0.034	0.029	0.024
Cys40I	116	117	127	121	0.010	0.004	0.005	0.023	0.015	0.013
Mean (σ)	120 (2.9)	116 (1.2)	125 (2.2)	121 (1.1)	0.0205 (0.011)	0.014 (0.013)	0.005 (0.003)	0.030 (0.015)	0.025 (0.013)	
rms with Leu38I					0.023	0.018	0.005	0.033	0.027	
rms without Leu38I					0.025	0.020	0.006	0.037	0.030	

model. As a result, this group could shift to absorb any errors in the model or in the data.

D. Solvent Structure

Systematic searches of difference maps (program of J. Moult, modified by M. Fujinaga) at various stages of refinement led to the inclusion of 179 water molecules in the present structure. Peak searches were conducted until the heights of the remaining peaks were less than 3σ above the rms error of the map. The rms of the map was calculated to be 0.06 $e^-/Å^3$ (program of J. Moult) for 8.0 to 2.1Å data. Initially, the number of peaks assigned to water molecules was 210. At the end of the refinement, the waters were removed 30 at a time, the model refined for a few cycles, and a difference map calculated. Eliminated waters that did not reappear on the difference map or that appeared as very small peaks (less than two grid points) were not included in the final model. This method is similar to one employed during the refinement of the structure of α-lytic proteinase (Fujinaga *et al.*, 1985).

During the course of the refinement of the enzyme-inhibitor complex, three strong peaks in the electron density map were modeled as solvent ions. Two of these peaks were modelled as sulphate anions, though these could also be phosphates (Figure III.10). One of the anions (#1) has only one full counterion as a ligand (His158); the remaining charge balance may come from the hydrogen atom donor ligands. A more striking example of this is given by the sulphate binding protein (Quiocho *et al.*, 1987), where there are no fully charged ligands in direct contact with the sulphate anion. On the other hand, the other anion (#2), has two cationic ligands (Arg139 and Arg182), in addition to numerous hydrogen donating ligands (Figure III.10). It should be noted, however, that Arg 139 N^{η_2} forms an ion pair with Asp31 O^{δ_1} (2.9Å). Thus, it may be that Arg139 should not be considered a fully charged ligand with respect to the sulphate anion. Despite the differences in the coordination geometry of the two anions, both have similar B-factors: for sulphate 1, the values range from 30 to 32$Å^2$, and for sulphate 2 the range is 28 to 32$Å^2$. The ligand to

#1

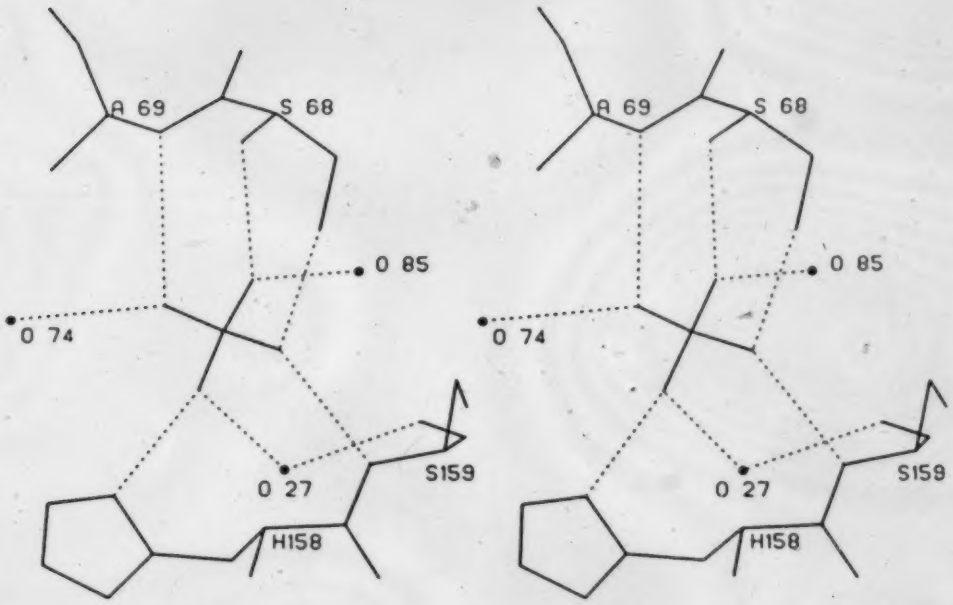
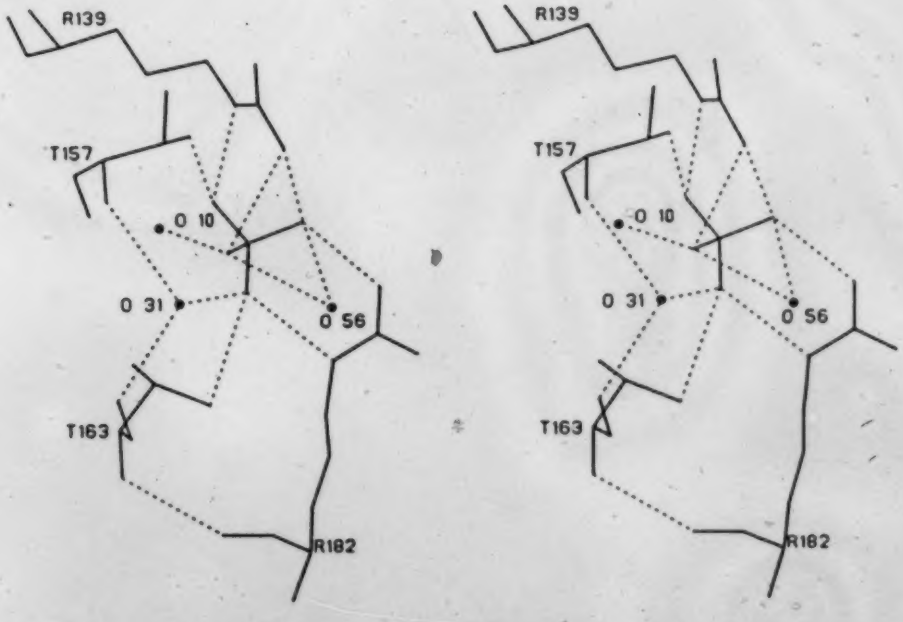


Figure III.10. Stereo figures showing the environments of the two sulphate/phosphate ions. Ligand to ion distances are given in Table III.VI. Dashed lines are hydrogen bond interactions, and filled circles are water molecules. The anions were refined as sulphate ions, with sulphur to oxygen distances of 1.5Å and oxygen to oxygen distances of 2.45Å.

#2



anion distances are given in Table III.VI.

The third peak was ascribed to a potassium/calcium ion which seems to play a key role in crystal packing. This cation is chelated by ligands from two symmetry related complexes (Figure III.11). The ligand to cation distances and the coordination angles are given in Table III.VII. The peak was modelled as a potassium cation since this appeared the only possibility among the cations added to the crystallization medium (NH_4^+ and K^+). The coordination geometry is pentagonal bipyramidal, which has been observed for potassium ions (Albertsson *et al.*, 1973), but the ion to oxygen distances are significantly shorter than those typically reported for K^+ to oxygen distances in small molecule crystal structures (2.6 to 3.0Å; Mariezcurrena and Rasmussen, 1973; Berman *et al.*, 1973). The distances are consistant with average distances (2.4Å) observed between water molecules and Ca^{++} (Einspahr and Bugg, 1980), and between carboxylate groups and Ca^{++} (Einspahr and Bugg, 1981) in small molecule crystal structures and in protein structures (Einspahr and Bugg, 1984). Furthermore, the mean distance is comparable to those observed in the refined structures of Troponin-C (2.3Å, Herzberg and James, 1985), and *S. griseus* trypsin (2.37Å, Read and James, 1988) for the calcium binding sites. Since K^+ and Ca^{++} have similar scattering properties, these distances provide strong evidence that the ion is, in fact, a calcium ion, despite the fact that calcium was not explicitly added to the medium. The resulting charge balance, then, would be even, with three negative carboxylate groups (Asp120K, Tyr243, and Arg51I), and three positive charges (Ca^{++}, and the guanidinium group of Arg51I).

The reactive site region of the complex with those waters that are within 5Å of both PCI-1 and SGPB is presented in Figure III.12. Though extensive bridging networks of solvent molecules exist between the polar groups of both proteins, the importance of these networks in inhibitor binding is not easily assessed. Thus, it would be difficult to predict any changes the equilibrium binding constant might incur as a result of side chain modifications that would disrupt these networks. Furthermore, changes that might alter the association

Table III.VI: Ligand-Sulphate Distances (Å)

Sulphate 1

Ligand		Sulphate Oxygen	
Atom	Symmetry		
Ser 68 Oγ	x,y,z	O1	2.6
Ser 159 N	1-x,y+1/2,1-z	O1	2.8
Wat 74	x,y,z	O2	3.0
Ala 69 N	x,y,z	O2	3.6
Wat 27	1-x,y+1/2,1-z	O3	2.8
His 158 Nδ1	1-x,y+1/2,1-z	O3	2.8
Wat 85	x,y,z	O4	2.6
Ser 68 N	x,y,z	O4	2.9

Sulphate 2

Ligand		Sulphate Oxygen	
Atom	Symmetry		
Wat 31	1-x,y+1/2,1-z	O1	2.7
Thr 163 Oγ	1-x,y+1/2,1-z	O1	3.1
Arg 182 Nε	1-x,y+1/2,1-z	O1	2.8
Wat 10	x,y,z	O2	2.6
Arg 139 Nη1	x,y,z	O2	3.3
Wat 56	1-x,y+1/2,1-z	O2	3.2
Arg 182 Nη2	1-x,y+1/2,1-z	O3	2.7
Wat 56	1-x,y+1/2,1-z	O3	3.4
Arg 139 Nη1	x,y,z	O3	3.4
Thr 157 Oγ	x,y,z	O4	2.6
Arg 139 Nε	x,y,z	O4	3.1

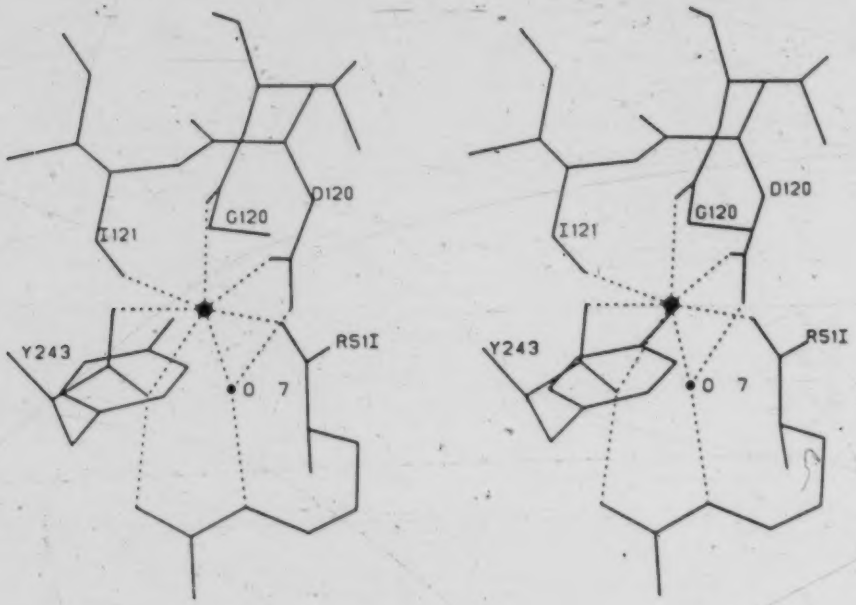


Figure III.11. Stereo representation of the environment of the Ca^{++}/K^{+} ion. Ligand to ion distances and angles are given in Table III.VII. Smaller filled circle is a water molecule. Additional interactions are: Arg51I $N^{\eta 1}$ to Tyr243 O, 2.6Å, Arg51I N^{ϵ} to Wat 7, 2.7Å, and Asp120K $O^{\delta 1}$ to Wat 7, 2.7Å.

constant of an inhibitor with a given enzyme could not be expected to give the same results with different enzymes. Similar considerations are discussed in the comparison of the structure of OMTKY3 bound to two different enzymes, SGPB and α-chymotrypsin (α-CHT; Fujinaga *et al.*, 1987).

E. Comparison of PCI-1 to Other Solved Inhibitors

The fold of PCI-1 is novel compared to all previously determined inhibitor structures. In spite of this, the reactive site loop of PCI-1 is conformationaly equivalent to those of the other inhibitor families.

The summary of contacts between PCI-1 and SGPB, presented in Table III.IV, is similar to results obtained for other inhibitor-enzyme complexes. As in the OMTKY3-SGPB complex (Read *et al.*, 1983), the OMTKY3-α-CHT complex (Fujinaga *et al.*, 1987), the eglin-c-subtlisin Carlesberg complex (McPhalen, 1986) and the CI-2-subtilisin Novo complex

Table III.VII: Ligand-Ion Distances and Coordination Angles for K^+/Ca^{++} Ion

Atom	Symmetry	Distance (Å)
Arg 51I O	x,y,z	2.2
Tyr 243 O	x,y,z+1	2.3
Tyr 243 OT	x,y,z+1	2.5
Asp 120K $O^{\delta 2}$	x,y,z+1	2.3
Ile 121 O	x,y,z+1	2.4
Gly 120 O	x,y,z+1	2.5
Wat 7	x,y,z+1	2.5
Mean		2.38

Atom 1*	Atom 3	Angle (°)
Tyr 243 O	Wat 7	74
Tyr 243 O	Tyr 243 OT	55
Tyr 243 OT	Gly 120 O	78
Gly 120	Asp 120K $O^{\delta 2}$	78
Asp 120K $O^{\delta 2}$	Wat 7	76
Ile 121 O	Tyr 243 OT	80
Ile 121 O	Tyr 243 O	91
Ile 121 O	Wat 7	87
Ile 121 O	Asp 120K $O^{\delta 2}$	86
Ile 121 O	Gly 120 O	87
Arg 51I O	Tyr 243 O	97
Arg 51I O	Tyr 243 OT	101
Arg 51I O	Wat 7	100
Arg 51I O	Asp 120K $O^{\delta 2}$	91
Arg 51I O	Gly 120 O	84
Arg 51I O	Ile 121 O	171

* In each case, the apex of the angle (Atom 2) is the ion.

(Mcphalen *et al.*, 1985b), the P_1 residue of PCI-1 makes the most contacts of any single residue. Asp41I makes no contacts to SGPB, unlike the P_1' residue in the other complexes. Rather, it faces the main body of the inhibitor and is involved in an Asx-type turn (see Table III.III).

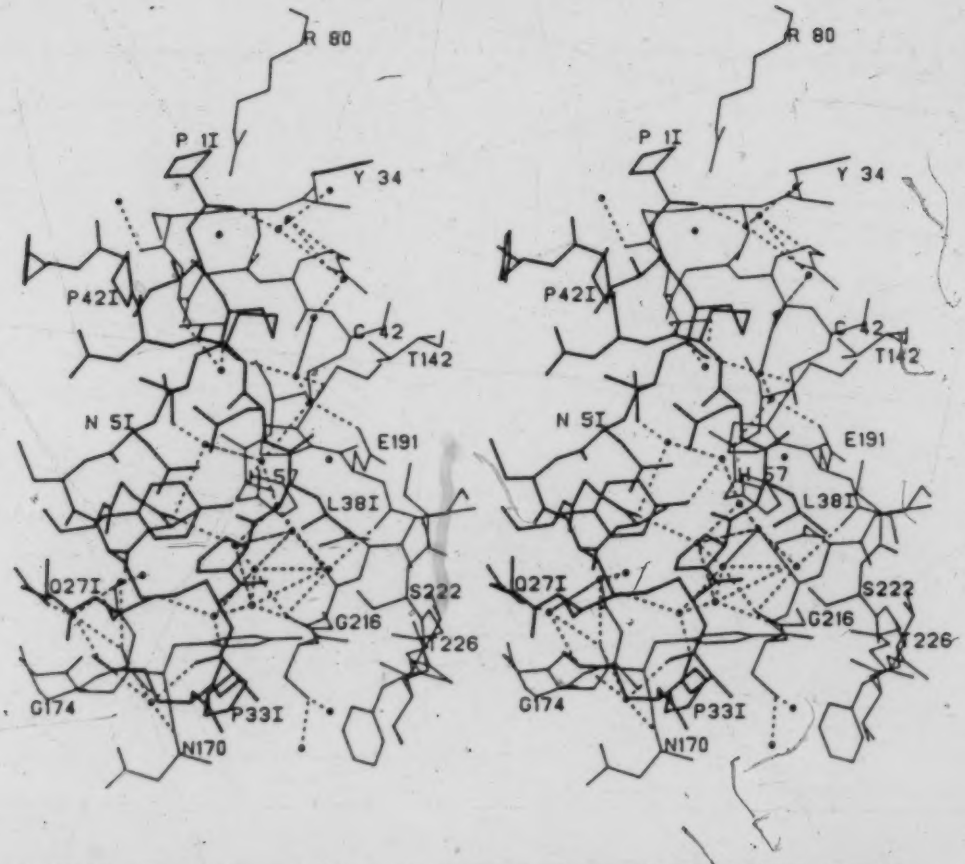


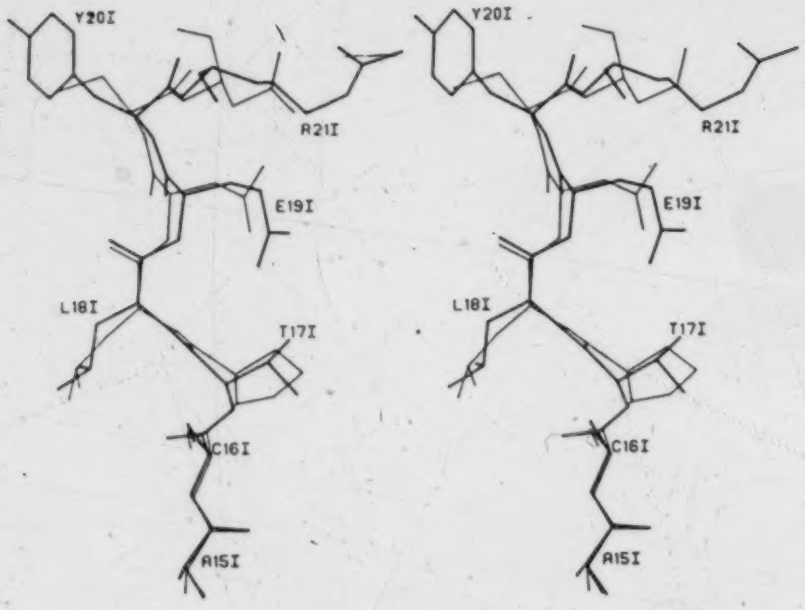
Figure III.12. Stereo figure of solvents in reactive site of PCI-1·SGPB complex. PCI-1 is drawn in thick lines, SGPB in thin lines, and hydrogen bonds are drawn dashed. Water molecules, shown as filled circles, were included only if they were within 5.0Å of any atom of both PCI-1 and SGPB. The solvent molecule forming a hydrogen bond to Ala190 O is conserved in the structure of OMTKY3·SGPB complex, and in the structure of SGPA (Sielecki *et al.*, 1979), as discussed in the text.

A least squares overlap procedure (program of W. Bennett) was used to compare the P_4-P', main chain atoms of CI-2, eglin-c, and OMTKY3 (in complex with both SGPB and α-CHT) with those of PCI-1. In addition to the main chain atoms, the C^{β} of the P_1 residue of all the inhibitors was included in the overlap. The root mean square (rms) distance of each of the reactive sites from PCI-1 was: 0.38Å for eglin-c, 0.43Å for CI-2, 0.49Å for OMTKY3 (SGPB), 0.46Å for OMTKY3 (α-CHT), and 0.75Å for PTI. The carbonyl oxygen atom of P', was excluded from the least-squares fit calculations because of large deviations (>2.0Å for OMTKY3 and eglin-c) relative to PCI-1. The extent to which

the reactive sites assume closely related conformations can be seen in Figure III.13, where PCI-1 is compared to four other inhibitors. This relationship is also evident in Table III.VIII, which compares the (ϕ,ψ) values for the P_4-P'_3 residues of the inhibitors shown in Figure III.13; the values for OMTKY3 bound to α-chymotrypsin are included, though this molecule is not shown in the figure.

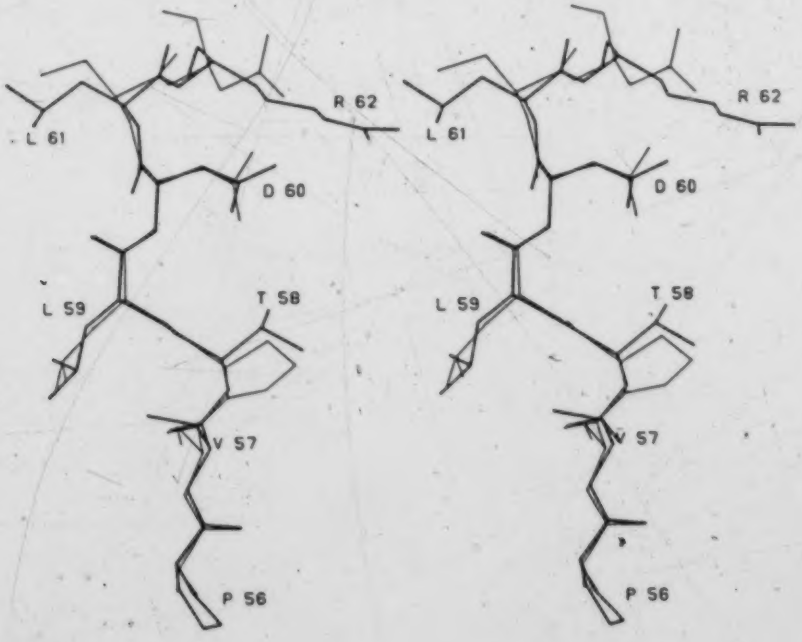
The geometry of the P_1 leucine side-chain in PCI-1 is significantly distorted. The chiral volume at the C^{α} was 2.23Å^3 (ideal is 2.49Å^3), and at the C^{γ} was -1.92Å^3 (ideal is -2.68Å^3), the largest deviation of the whole complex. Though two different conformations were modelled, peak searches consistantly showed unaccounted density at the location of the C^{γ} of the alternate conformation. These observations strongly suggest that there are two conformers of Leu38I present in the crystals of complex, and the distortion is an attempt to model an average structure. This need not imply rotation of the Leu38I side chain in the complex; rather, selection of a given conformation could occur upon binding.

The P'_1 residue of PCI-1 is Asn39I. CI-2 and OMTKY3 have a glutamate at P'_1 and eglin-c has an aspartate. It has been postulated (Fujinaga *et al.*, 1982) that formation of a hydrogen bond by the side chain of the P'_1 residue with its own main-chain nitrogen atom should raise the activation energy for nucleophilic attack at the P_1 carbonyl carbon atom. In addition, *ab initio* quantum mechanically derived electron densities (Moult and James, unpublished results) show that the interaction of a negatively charged carboxylate oxygen bonded to the amide hydrogen atom of the scissile peptide increases the delocalization in the carbonyl bond, thus increasing the stability of the peptide link. Although Asn39I of PCI-1 does not have a negative charge, the corresponding hydrogen bond to $O^{\delta 1}$ is formed. It should be noted that although PTI has an alanine as the P'_1 residue, there is an analogous hydrogen bond to the carbonyl oxygen of Gly36 in the inhibitor (Huber and Bode, 1978). Furthermore, in the recently determined sequences of 100 avian ovomucoid inhibitors (Laskowski *et al.*, 1987), all but one have a glutamate or an aspartate, at the P'_1 position. The exception is the inhibitor from the road runner, which has a leucine in the P'_1 position.

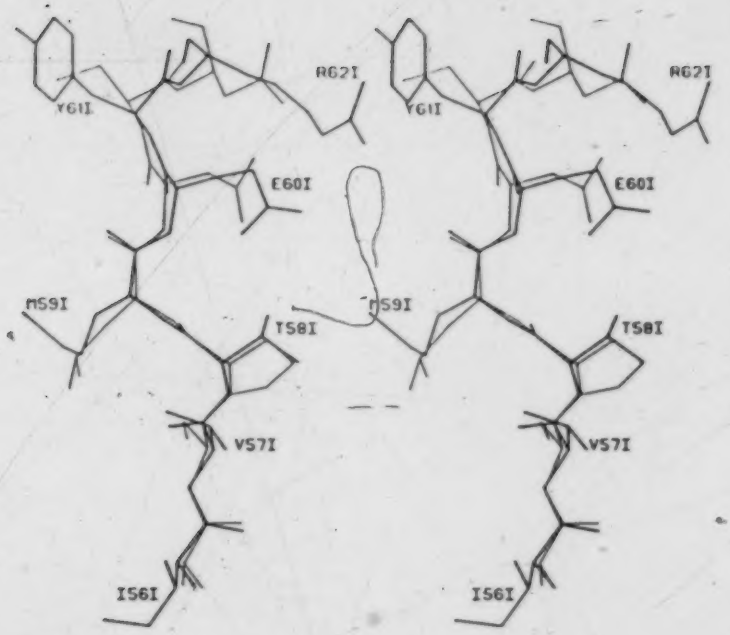


A

Figure III.13. Stereo representations of overlaps of OMTKY3, eglin-c, CI-2, and PTI to PCI-1, by the method described in the text. PCI-1 is drawn with thin lines in each diagram. (A) OMTKY3. (B) Eglin-c. (C) CI-2. (D) PTI.

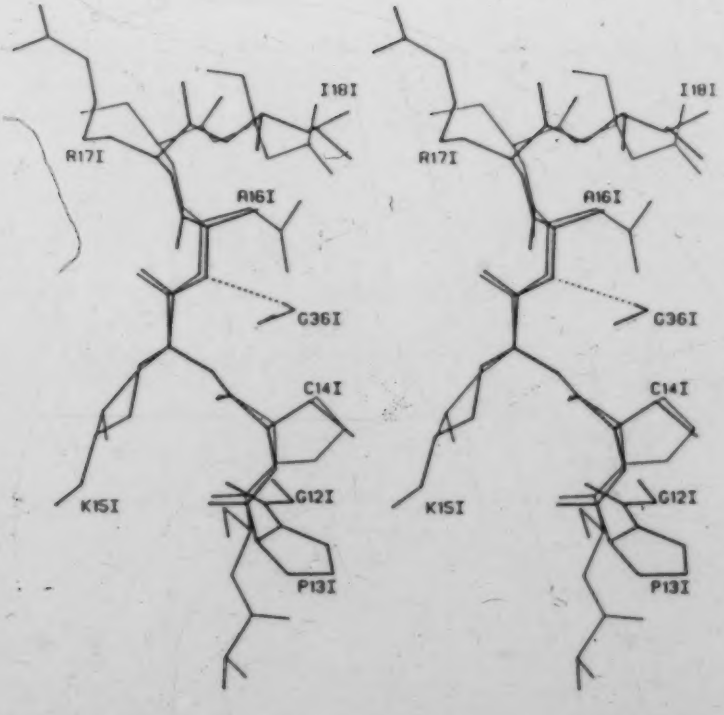


B



C

Figure III.13 (cont'd)



D

Table III.VIII: Comparison of main-chain conformational angles from several inhibitors

	P_4		P_3		P_2		P_1		P'_1		P'_2		P'_3	
	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
PCI-1	-82	132	-137	152	-65	151	-115	51	-102	165	-95	121	-71	132
OMTKY3 (SGPB)	-158	158	-126	147	-69	162	-119	45	-84	155	-100	115	-147	92
OMTKY3 (α-Chymo.)	-129	136	-131	150	-68	160	-107	32	-74	159	-113	107	-141	76
CI-2	-93	140	-133	166	-64	147	-103	34	-91	146	-106	119	-118	113
Eglin-c	-76	138	-143	165	-65	151	-112	42	-96	176	-120	113	-119	115
PTI	78	175	-77	-30	-70	156	-117	39	-88	164	-112	80	-106	122

The effect of this substitution was not reported.

The complete explanation for the size of the barrier to hydrolysis cannot be found in the structure of the inhibitor alone. Studies with OMTKY3 and several different enzymes (Ardelt and Laskowski, 1985) and one involving PTI with α-chymotrypsin and trypsin (Quast *et al.*, 1978) have shown that the ratios of the off rates for the virgin inhibitor to the off rates for the cleaved or modified inhibitor vary greatly depending upon the enzyme used. Thus any explanation for the variation in these ratios, which differ by up to six orders of magnitude, must include the structure of the enzyme and indeed the dynamic aspects of the molecular complexes themselves (Read and James, 1986).

F. Comparison of the SGPB structures from the PCI-1·SGPB and OMTKY3·SGPB complexes

A least-squares overlap (program of W. Bennett) of the main-chain atoms of SGPB from the two complexes gave an rms deviation of 0.33Å for 735 of the 741 main-chain atoms in SGPB. The overlap program has a feature for omitting from the calculation those atoms that deviate by more than a certain cutoff value (in the present case, 1.9Å). Only one section of chain had deviations greater than this value: the loop which includes Ser38 and Thr39. The maximum deviation for this section of chain was 3.7Å for Thr39 O. The difference in the conformations of the loops can be seen in Figure III.14. If the loop maintained the same conformation as in the OMTKY3·SGPB complex, the carbonyl carbon of Ser38 would make a close contact of 1.3Å with Pro42I C^{β} of PCI-1. In the PCI-1·SGPB complex, Thr39 O^{γ} takes the place of a water molecule in the OMTKY3·SGPB complex in making hydrogen bonds to Cys58 O and Ser35 O^{γ}. The B factors for the main-chain atoms of this loop in the PCI-1 complex lie between 20.0Å^2 (Ser38 C^{α}) and 20.7Å^2 (Thr39 O), while in OMTKY3 complex the range is from 20.2 (Ser38 N) to 24.8Å^2 (Thr39 O).

The section of chain from Gly172 to Gly174 had main-chain atom deviations of greater than 1.0Å, with a maximum value of 1.4Å for the carbonyl oxygen of Gly173. In the PCI-1 complex the B-factors of the main-chain atoms of this loop are between 18.5 and

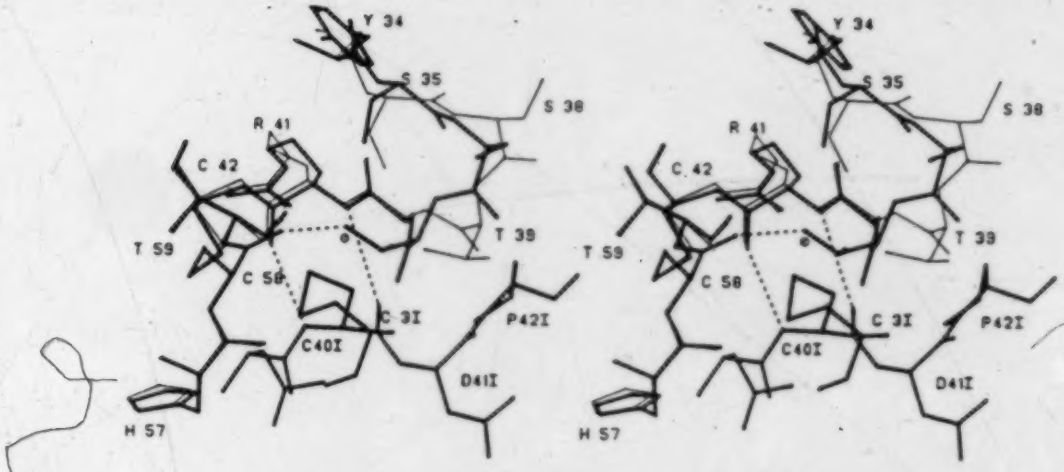


Figure III.14. Stereo plot of the Ser35 to Gly40 loop of SGPB showing the different conformations assumed in the complexes with PCI-1 (thick lines) and OMTKY3 (thin lines). Dashed lines are hydrogen bond interactions in the PCI-1 complex. Additional interactions are found between Thr39 O^{γ} and Ser35 O^{γ} (2.9Å), Thr39 NH and Ser35 O^{γ} (3.0Å), Gly40 NH and Ser35 O (3.1Å), and Ser35 NH and Gly40 O (3.0Å). The interactions shown have the following distances: Cys40I NH to Arg41 O, 3.1Å, Arg41I NH to Cys40I O, 3.4Å, and Thr39 O^{γ} to Cys58 O, 2.6Å. The open circle is a solvent molecule from the OMTKY3 complex which is 0.3Å from O^{γ} of Thr39 of the PCI-1 complex, based on the overlap described in the text.

20.1Å^2. In the OMTKY3 complex, the B-factors range from 19.0Å^2 for Gly172 N to 32.2Å^2 for Gly173 O. The lowering of the thermal motion parameters of the residues in the PCI-1·SGPB complex is most probably due to crystal contacts to a symmetry related PCI-1 molecule at (-x,-1/2+y,1-z). This loop can be seen in Figure III.15.

A comparison of the average main chain B-factors for SGPB from the two complexes (Figure III.16) shows that there are only a few regions where there are large differences. If one considers those regions which have $|\Delta\langle B\rangle|$ of 5.0Å^2 or greater, then these differences can be rationalized in terms of differences in crystal packing. The environments of each these regions in the two complexes are summarized in Table III.IX, which shows the apparent correlation between the number of contacts and the relative magnitude of the main-chain $\langle B\rangle$.

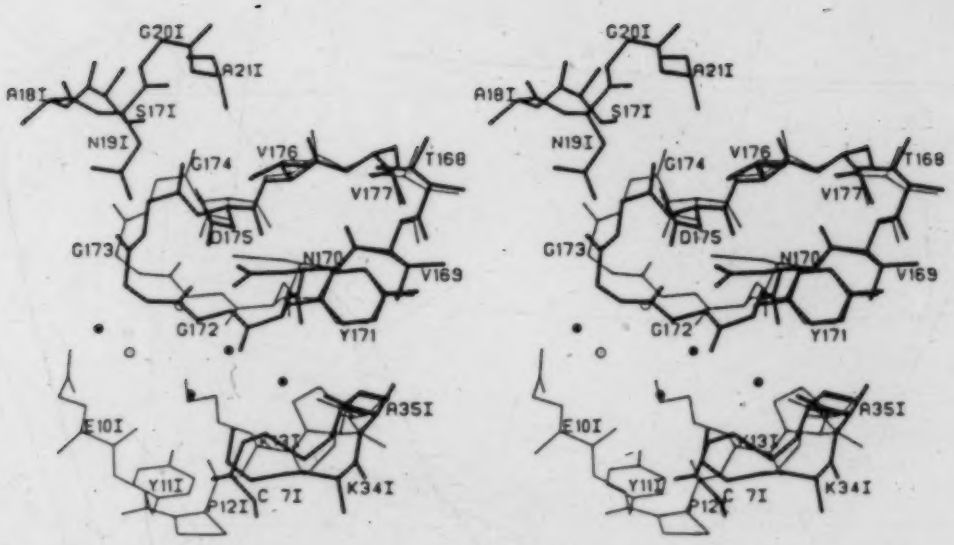


Figure III.15. Stereo drawing showing difference in conformation between Gly172 and Asp175 of SGPB in two complexes, PCI-1·SGPB (thick lines) and OMTKY3·SGPB (thin lines). Relevant portions of PCI-1 and OMTKY3 are also shown in thick and thin lines, respectively. Ordered waters are also shown as filled (PCI-1) and open (OMTKY3) circles. The segment of PCI-1 from Ser17I to Ala21I is from a symmetry related molecule (-x,-1/2+y,1-z). The distance from Ser17I O^{γ} to the carbonyl oxygen of Gly174 is 2.5Å.

One additional region, involving Gly59B, has significant $\Delta\langle B\rangle$ values arising from causes other than crystal packing. The probable cause of the reduction in the average B-factors for these residues in the PCI-1 complex, relative to the OMTKY3 complex, is the closer proximity of Ser38 and Thr39 in the former complex (*vide supra*). There are a total of 8 contacts between the segment Gly59B-Ala60 and Ser38-Thr39, the shortest being Ala60 N to Thr39 $C^{\gamma 2}$ (3.5Å).

A similar comparison between the main-chain $\langle B\rangle$ values of SGPB from the complex with PCI-1 and native SGPB is shown in Figure III.17. In this case, differences in main-chain B factors arise from the fact that an inhibitor is bound in the former case and not in the latter, and from crystal packing. Contacts to the inhibitor have been summarized in Table III.IV. The anomalous decrease in mobility of the region Gly172-Gly174 in the native enzyme can be attributed to the large number of symmetry contacts which occur in

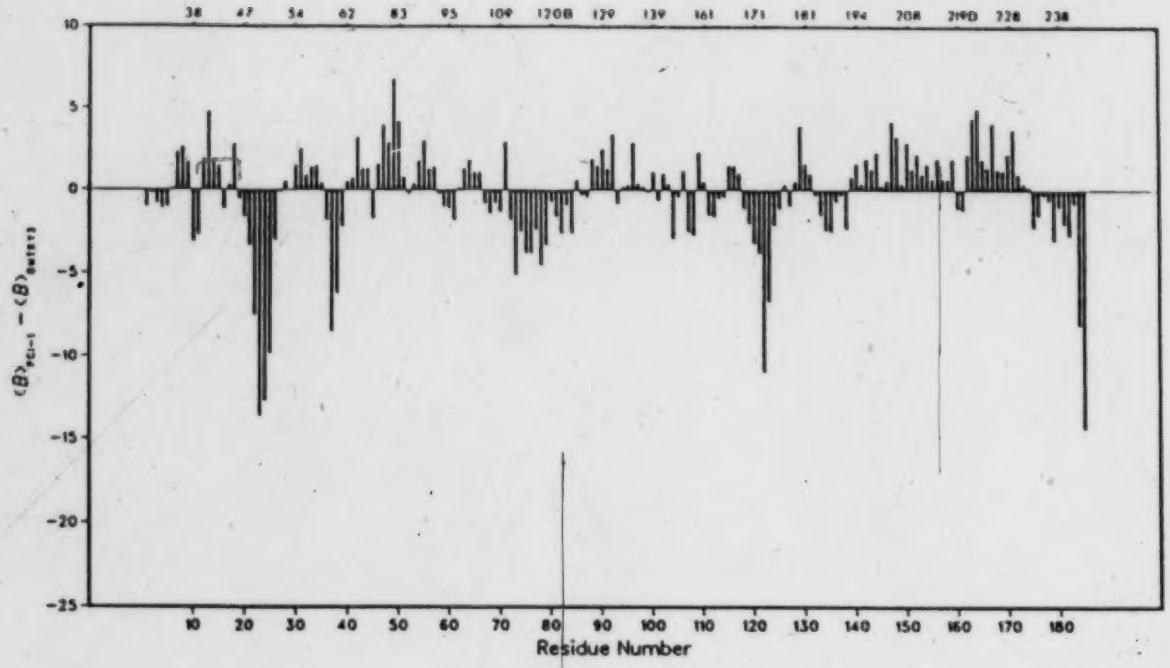


Figure III.16. Bar graph comparing the average main-chain B-factors of SGPB in the two complexes with PCI-1 and OMTKY3. The sequential numbering of SGPB appears at the bottom of the horizontal axis, while numbering based on the alignment with α-chymotrypsin appears at the top. The vertical axis represents $\Delta\langle B\rangle$, and the units are $Å^2$.

Table III.IX: Summary of symmetry contacts in crystals of PCI-1·SGPB and OMTKY3·SGPB

Region	PCI-1·SGPB	OMTKY3·SGPB
Ser48A-Thr49		
range of <B>[a]	13-19 Å^2	22-32Å^2
# of general symmetry contacts[b]	12	6
hydrogen bonds (dist.) (symmetry)[b]	Gly23 O-Asn14I N$^{\delta 2}$ (3.1Å) (x,y,z-1) Ser48 C-Tyr16I O$^{\eta}$ (2.6Å) (x,y,z-1)	(none)
Thr82		
range of <B>	23Å^2	16Å^2
# of general symmetry contacts	0	1
hydrogen bonds (dist.) (symmetry)	(none)	(none)
Gly173-Gly174		
range of <B>	19-20Å^2	23-31Å^2
# of general symmetry contacts	6	0
hydrogen bonds (dist.) (symmetry)	Gly173 O-Asn19I N$^{\delta 1}$ (3.1Å) (-x,y-1/2,-z+1) Gly174 O-Ser17I O$^{\gamma}$ (2.5Å) (-x,y-1/2,-z+1)	(none)
Val242-Tyr243		
range of <B>	13-14Å^2	22-28Å^2
# of general symmetry contacts	18	0
hydrogen bonds (dist.) (symmetry)	Tyr243 O-Arg51I N$^{\eta 1}$ (2.7Å) (x,y,z-1)	(none)

[a] Main-chain atoms only
[b] These are only contacts to other protein atoms, not to symmetry generated solvents

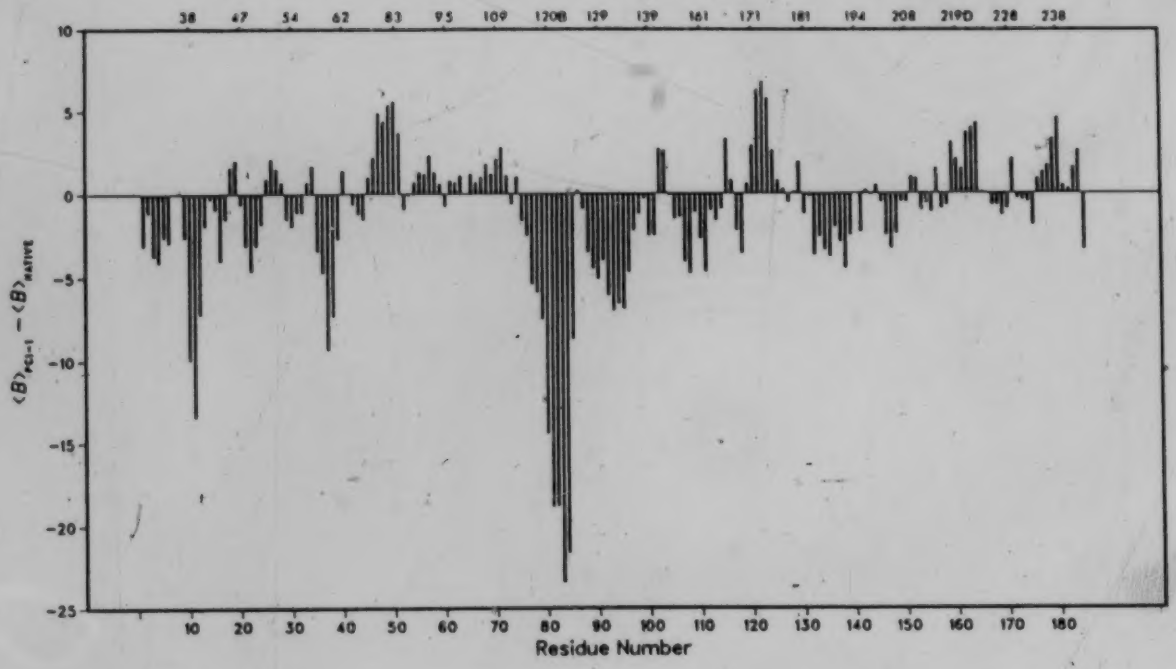


Figure III.17. Graph similar to Figure III.16, but comparing the average main-chain B-factors for SGPB in complex with PCI-1 to those values from native SGPB.

crystals of the free enzyme. Though there are symmetry contacts to this region in the crystals of the complex, these are not as numerous as the contacts which occur in the crystal of the native enzyme. This is summarized in Table III.X, along with data for other regions of SGPB where changes in main-chain <B> values are caused by differences in crystal packing. The cause of the significant Δ<B> values for Gly59B and Ala60 are the same as those given above in the comparison between SGPB bound to PCI-1 and OMTKY3.

G. Comparison of Solvent Structure in the Two Complexes

To compare the solvent structures of the two complexes, the following procedure was used. The X, Y, and Z limits of the PCI-1·SGPB complex, the OMTKY3·SGPB complex, and the structure of native SGPB were extended by 10Å to define boxes which were filled with all symmetry related solvent molecules generated from unique sets. The three solvent sets were then superimposed based on the overlaps of the main-chain atoms of SGPB, using the enzyme from the complex with PCI-1 as the fixed molecule. The solvent set from the native structure (1126 atoms) was used as the standard against which the other two solvent boxes were compared. Two solvent molecules were considered to be "identical" if they were within 1.0Å of each other in the two sets. The PCI-1·SGPB solvent set (861 atoms) had 97 solvent molecules in common, while the OMTKY3·SGPB solvent set (966 atoms) had 92 solvent molecules in common with the native set. Atoms found in both lists (47 atoms) can be considered common to all three forms of SGPB. From this list of 47 atoms, however, only two were within 5.0Å of both the enzyme and PCI-1, when the solvent coordinates from the PCI-1·SGPB complex were used. One molecule forms a hydrogen bond with Gly219D NH (2.6Å), and the other forms hydrogen bonds with Ala190 O (2.7Å), and Gly219D O (2.8Å). Neither molecule forms any hydrogen bonds to the inhibitor (Figure III.12).

This result implies that the solvent structure in the region of the active site of the native enzyme is almost entirely replaced or displaced when the inhibitor binds. Thus, it

Table III.X: Summary of symmetry contacts in crystals of PCI-1·SGPB and native SGPB

Region	PCI-1·SGPB	Native SGPB
Thr82-Val83		
range of <B>[a]	20-23Å^2	15-17Å^2
# of general symmetry contacts[b]	0	3
hydrogen bonds (dist.) (symmetry)[b]	(none)	(none)
Asp119-Ile121[c]		
range of <B>	9.3-22Å^2	16-41Å^2
# of general symmetry contacts	7	0
hydrogen bonds (dist.) (symmetry)	Gly120 N-Arg51I O (3.2Å) (x,y,z-1) Gly130 N-Arg51I OT (2.8Å) (x,y,z-1)	(none)
Asn129-Met134		
range of <B>	7.1-14Å^2	14-20Å^2
# of general symmetry contacts	6	4
hydrogen bonds (dist.) (symmetry)	(none)	(none)
Gly172-Gly174		
range of <B>	19-20Å^2	13-14Å^2
# of general symmetry contacts	6	25
hydrogen bonds (dist.) (symmetry)	Gly173 O-Asn19I $N^{\delta 1}$ (3.1Å) (-x,y-1/2,-z+1) Gly174 O-Ser17I O^{γ} (2.5Å) (-x,y,-1/2,-z+1)	Gly173 N - Val184 O (2.9Å) (x,y,z+1)

[a] Main-chain atoms only

[b] These are only contacts to other protein atoms, not to symmetry generated solvents

[c] These residues are involved in chelating to Ca^{++}/K^{+} atom in the crystals of PCI-1·SGPB; see Table III.VII and Figure III.9.

appears that the final solvent structure around the enzyme active site in these complexes is mostly dependant upon the combined topology of the two molecules, rather than that of the enzyme.

One of the two conserved water molecules that was found contacting both the inhibitor and the enzyme in the two complexes makes a total of three contacts to the side chain of Leu38I (Figure III.18). The shortest contact is 3.6Å to $C^{\delta 1}$. The molecule makes 20 contacts to SGPB including the two hydrogen bonds. An equivalent water molecule is also found in the refined structure of SGPA (Sielecki *et al.*, 1979). It could be argued that replacement of the side chain of Leu38I with one containing an unbranched aliphatic chain with a terminal hydroxyl group could displace this water. This would likely increase the association constant, and so render the inhibitor even more effective. During attempts to

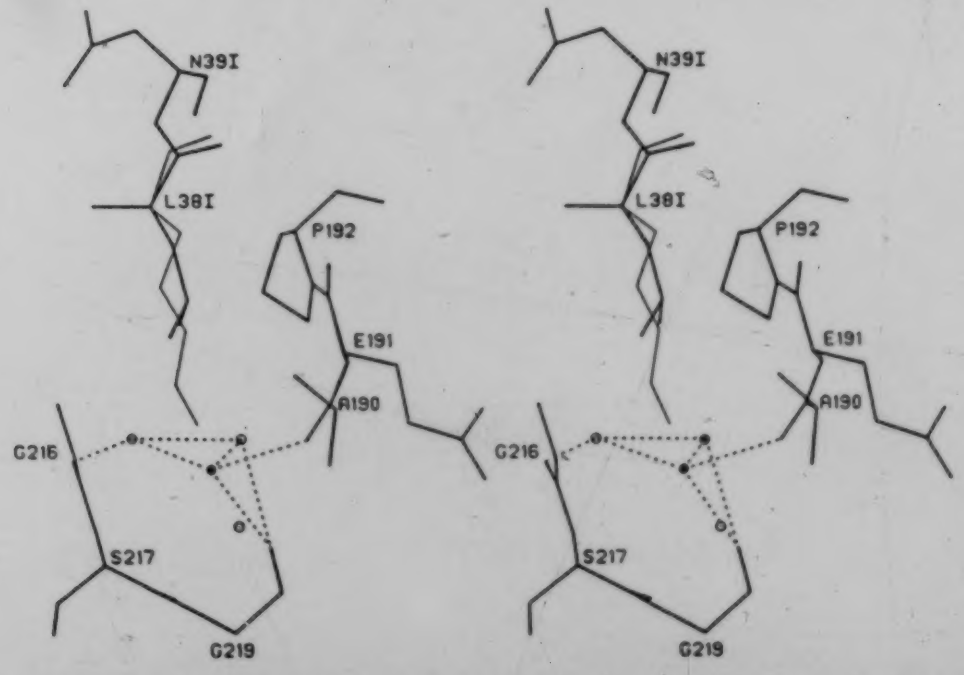


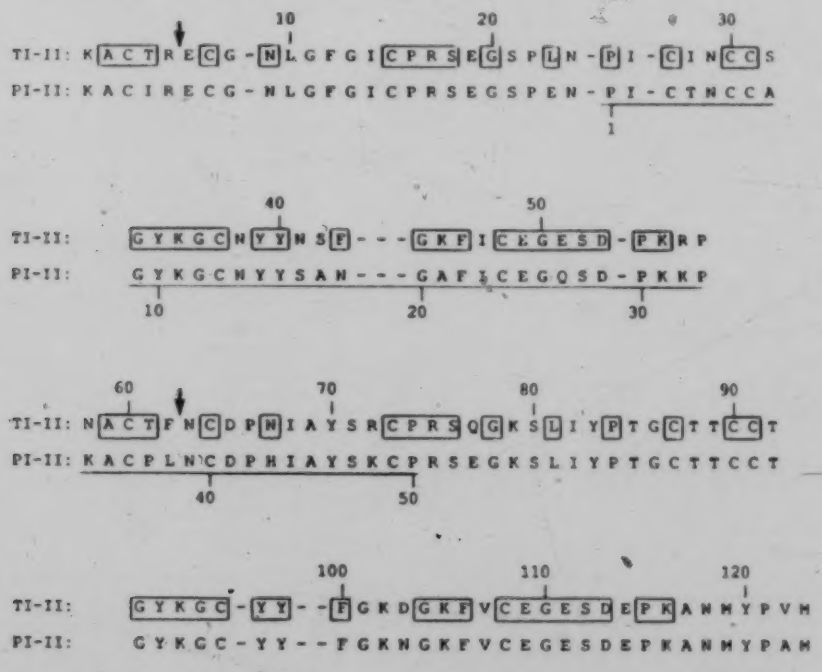
Figure III.18. Stereo plot showing one of two solvent molecules (filled circle) in contact with both the inhibitor and the enzyme which is conserved between the three forms of SGPB (native, and in complex with PCI-1 and OMTKY3). Displacement of this molecule with an hydroxyl group attaced to the P_1 residue of PCI-1 might increase the association constant of the inhibitor. An attempt to model this is shown using a lysine side chain (thin lines). The distance of the terminal atom of the side chain to the water molecule in question is ~1.0Å.

model such a side chain in the active site of SGPB using a lysine residue, it quickly became apparent that there were many water molecules which could be displaced. Furthermore, several of these alternatives did not require the unfavourable torsion angles required to bring the proposed hydroxyl group within 1Å of the target solvent molecule (Figure III.18). The model shown in the figure is based on results obtained with SGPA and peptide amide substrates (James *et al.*, 1980). In these studies, the solvent molecule in question was displaced by the *p*-hydroxyl group of a P_1 tyrosine reidue of a bound tetrapeptide product complex. The x^1 values for the tyrosine residue was -94° and for the lysine side chain in Figure 19, x^1 is -52°.

H. Relationship of PCI-1 to Inhibitor II

Inhibitor II, isolated from either tomatoes or potatoes, inhibits both chymotrypsin and trypsin. The c-DNA sequence (Thornbrug *et al.*, 1987), and the derived amino-acid sequence of these two inhibitors (Table III.XI) shows the two domain organization of these two molecules clearly, and provides an explanation for this inhibitory property. The boxed regions on tomato Inhibitor II indicate that the sequence of the protein is made up of two repeats, with each repeat containing one reactive loop, located near the N-terminal of the repeat. The amino terminal repeat contains the trypsin specific region, and the carboxy terminal repeat contains the chymotrypsin specific region. In Table III.XI, the scissile bond for each region is indicated by an arrow. Southern blot analysis suggests that there is a gene family (8 to 10 copies) producing Inhibitor II proteins, all with slight variations in amino acid sequence (results not shown). This is consistant with earlier observations regarding electrophoretic and immunological differences in components of Inhibitor II isolates (Bryant *et al.*, 1976). In fact, isoforms of Inhibitor II have been isolated which have two trypsin specific domains, rather than one trypsin specific and one chymotrypsin specific domain (unpublished results).

Table III.XI: Sequences* of Inhibitor II from Tomatoes and Potatoes


TI-II: K A C T R E C G - N L G F G I C P R S E G S P L N - P I - C I N C C S
PI-II: K A C I R E C G - N L G F G I C P R S E G S P E N - P I - C T N C C A
TI-II: G Y K G C N Y Y N S F - - - G K F I C E G E S D - P K R P
PI-II: G Y K G C N Y Y S A N - - - G A F I C E G Q S D - P K K P
TI-II: N A C T F N C D P N I A Y S R C P R S Q G K S L I Y P T G C T T C C T
PI-II: K A C P L N C D P H I A Y S K C P R S E G K S L I Y P T G C T T C C T
TI-II: G Y K G C - Y Y - - F G K D G K F V C E G E S D E P K A N M Y P V M
PI-II: G Y K G C - Y Y - - F G K N G K F V C E G E S D E P K A N M Y P A M

* The hyphens used in the sequences are merely present to align the apparent repeats, and are not necessarily structural requirements. Also, the repeats are identified strictly by identity; larger areas of residue similarity are evident. (Adapted from Graham *et al.*, 1985)

The amino acid sequence of PCI-1 (Hass *et al.*, 1982) can be matched identically with a region of potato Inhibitor II. This is indicated in Table III.XI, where the numbering beneath the residues of Inhibitor II (underlined region) is the numbering of PCI-1. Thus PCI-1 could result from post-translational processing of Inhibitor II, in which the protein is cleaved leaving the carboxy terminal reactive site loop intact. To confirm this hypothesis, two tests were carried out (Thornburg, R. W., Pearce, G., Ryan, C. A., unpublished). The first test involved probing nascent proteins, translated from potato mRNA, with antibodies against Inhibitor II. These studies were unable to detect any smaller nascent proteins related to Inhibitor II. The second test involved performing Northern blots of tuber mRNA with the Inhibitor II cDNA probe (Graham *et al.*, 1985). No smaller mRNA's could be detected by

this method. These results strongly suggest that mRNA for PCI-1 does not exist.

This hypothesis implies a processing cleavage of Inhibitor II between Asn24 and Pro25, a heretofore unknown specificity. Though a P_1 asparagine-specific endopeptidase was postulated in the processing of the concanavalin A (ConA) precursor (Bowles *et al.*, 1986), the isolation of the enzyme was not reported. Furthermore, the cleavages associated with ConA processing do not involve a P'_1 proline. The difficulty of an Asn-Pro specific enzyme may be overcome by cleavage closer to the amino-terminal of Inhibitor II, for example at Arg17, 8 residues before the proline in question. Then non-specific aminopeptidase activity could trim the peptide to Pro25, to form the amino-terminal of PCI-1. Steric hindrance caused by the residues of PCI-1 might stop the aminopeptidase at this point. Processing at the C-terminus could be a trypsin specific cleavage at Arg75 (Arg51I of PCI-1).

Alternatively, non-enzymatic cleavage of Asp-Pro peptide bonds is known to occur more readily than with other peptide linkages (Landon, 1977; Piszkiewicz *et al.*, 1970). Strongly acidic conditions (pH $<$ 2.5) are required for these hydrolyses. Asp-Pro cleavage was also noted in the sequencing of rabbit skeletal muscle myosin light chain kinase during the decitraconylation steps using 9% formic acid (Takio *et al*, 1986). Acid catalyzed deamidation of glutamine and asparagine side chains is also well known. Thus, hydrolysis of the side chain amide in the presence of acid, followed by non-enzymatic cleavage of the resultant Asp-Pro bond is a possible mechanism for the formation of the amino terminus of PCI-1 from Inhibitor II. It is not apparent, however, that the acidity of the plant cell vacuoles, where PCI-1 is localized, is sufficient to duplicate the *in vitro* conditions described above.

However this processing is accomplished, it is unclear whether or not it occurs at the expense of the other reactive site on Inhibitor II. This other reactive site domain is not the source of the trypsin-directed single inhibitor, PoTI. PoTI probably originates from processing of a form of Inhibitor II with two trypsin-specific domains, where the carboxy terminal reactive site is preserved in the final product. Evidence for this comes from the

very high sequence identity between PoTI and PCI-1 (Hass *et al.*, 1982). Isoelectric focusing experiments performed on samples of PCI-1 showed at least 5 bands with pI's ranging from 4.0 to 8.6 (results not shown). A similar result was achieved with PoTI, though the pI range was smaller (4.0 to 6.5). These results are consistant with the hypothesis that PCI-1 and PoTI are products of the isoforms of Inhibitor II, but they do not clarify the fate of the other domain. The electrophoretically distinct forms of PCI-1 and PoTI may include molecules which are active remnants of the other domain of Inhibitor II.

It is remarkable that the sequence of PCI-1 corresponds to a region which lies *across* the two repeats of Inhibitor II, rather than being contained within one repeat or the other. This is contrary to the common assumption that large repeated sequences will fold into separate domains linked by a short stretch of polypeptide chain. The avian ovomucoid inhibitors, for example, consist of three tandemly repeated sequences, and each repeat gives rise to a separate structural domain (Laskowski and Kato, 1980). If this were the structural basis for Inhibitor II, then cleavages to form PCI-1 would have to take place inside each domain, while leaving the linker region intact. Furthermore, the disulphide bridges would have to be reduced, the remaining fragment refold, and the disulphide bridges reform, before PCI-1 could assume the structure observed in the complex with SGPB.

While this sequence of events is possible, we propose an alternate structure for Inhibitor II which would simplify the process required to generate PCI-1. The proposal is based on the assumption that no refolding of PCI-1 occurs. It is possible to achieve a structural model of Inhibitor II that has two domains related by a pseudo two-fold axis. To do this, the remaining portions of Inhibitor II (Lys1-Asn24, and Thr85-Lys116) must fold in a conformation similar to that seen for PCI-1. Thus, one must first establish that the other domain can assume a structure similar to PCI-1. An alignment of the sequence of PCI-1 with the missing segments of Inhibitor II is shown in Table III.XII; all of the gaps that are shown in Table III.XI have been eliminated, save two. There is an insertion (Pro66, Pro42I of PCI-1) between the residues corresponding to Gly8 and Asn9. The other insertion,

Table III.XII: Alignment of the Chymotrypsin Domain (PCI-1) of Inhibitor II with the Trypsin Domain

	25 (1)[a]							35 (11)										45 (21)							
Chymotrypsin domain	P	I	C	T	N	C	C	A	G	Y	K	G	C	N	Y	Y	S	A	N	G	A	F	I	C	E	G
Trypsin domain	T	G	C	T	T	C	C	T	G	Y	K	G	C	Y	Y	F	G	K	N	G	K	F	V	C	E	G
	85										95										105					

	55 (31)									65 (41)											75 (51)		
Chymotrypsin domain	Q	S	D	P	K	K	P	K	A	C	P	L	N	C	D	P	H	I	A	Y	S	K	C	P	R
Trypsin domain	E	S	D	E	P	K	A	K	A	C	I	R	E	C	G	-	N	L	G	F	G	I	C	P	R
	115					117	1										10						17		

[a] Sequence numbers in brackets are for PCI-1.

Glu114, did not require modelling (*vide infra*). A model for the region surrounding the deletion (Gly8-Asn9) was generated using the program TOM. A database of 17 proteins was searched to find a stretch of polypeptide chain with ends in similar conformations as the amino acids surrounding the deletion. A segment of amylase (Matsuura *et al.*, 1984) from Lys36 to Trp42 was fit and regularized. The side chains of the entire model were checked, and any apparent clashes relieved by manual adjustments to x-angles. To eliminate bad main-chain contacts resulting from the loop building, and any other unfavourable interactions, the model was then subjected to energy minimization using the GROMOS program (van Gunsteren and Berendsen, 1987; Version 4 used). The crystallographically derived model for PCI-1 was also energy minimized for comparison. A summary of the results are shown in Table III.XIII. Only two atoms in the PCI-1 structure had shifts larger than 1.0Å, and these were Lys32I O (1.2Å) and Pro50I O (1.2Å). The differences between the two models of the trypsin domain of Inhibitor II were larger, mostly due to shifts required to relieve a short contact between the carbonyl groups of Gly11 and Tyr99. Figure III.19 shows the energy minimized model for the N-terminal domain of Inhibitor II, in a similar orientation to PCI-1 in Figure III.5b. Note that, despite the deletion of a proline between Gly8 and Asn9, and the subsequent change in the chain conformation, the side chain of Leu10 has adopted a conformation which brings the two C^{δ} atoms into positions similar to those of the corresponding C^{γ} atoms of Ile44I in PCI-1. Figure III.20 is a detailed view of the deletion (Gly8-Asn9) with the corresponding region of PCI-1 shown for comparison.

The backbone hydrogen bonding network in the trypsin domain is very similar to that found in the chymotrypsin domain (PCI-1), if one excludes those hydrogen bonds involving side chains or involving the region around the deletion (Gly8-Phe12). Two main chain hydrogen bonds shown in the trypsin domain do not appear in PCI-1; these are Lys105 NH to Asn103 O, and Glu6 NH to Ile4 O. It should also be noted that the replacement of Asn5I of PCI-1 with Thr89 in the trypsin domain of Inhibitor II has removed, in this model,

Table III.XIII: Summary of *in-vacuo* [a] energy minimization results

	PCI-1		Inhibitor II trypsin domain	
	Initial	Final	Initial	Final
Total energy[b]	-394	-2190	2350	-1730
Change in total energy from cycle 990	-3		-2	
No. of non-bonded atom pairs	17,332	17,216	16,153	16,115
Bond energies				
not involving hydrogen	288	54.7	278	63.4
involving hydrogen	1.38	2.15	1.28	3.04
Bond angle energies				
not involving hydrogens	551	227	383	181
involving hydrogens	98.9	16.7	101	15.3
Improper dihedral energies				
not involving hydrogens	218	29.2	485	34.2
involving hydrogens	2.58	21.3	2.64	15.1
Dihedral energies				
not involving hydrogen	414	250	295	188
involving hydrogen	1.64	6.01	0.129	6.68
Third neighbour interactions				
electrostatic	-1040	-1390	-552	-1010
Lennard-Jones	-930	-1410	1360	-1230
Solvent accessible surface area[c] ($Å^2$)	3506	3620	3597	3646

[a] 37D4 force field used
[b] All energies are in KJ/mole
[c] Program of J. Moult

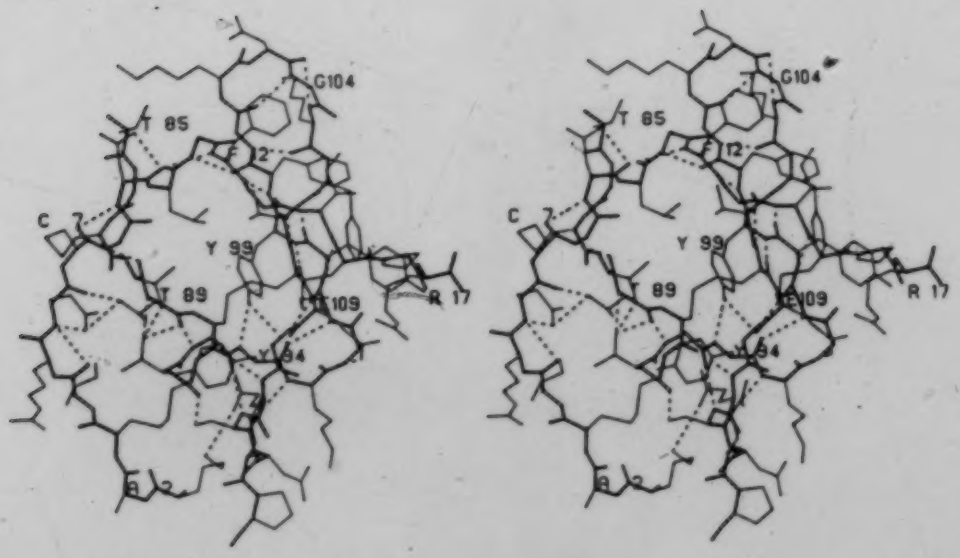


Figure III.19. Stereo figure of the energy minimized model of the trypsin specific domain of Inhibitor II. The residues from Lys116 to Met123 have not been included, since there is no corresponding segment of PCI-1 to use as a model. The conformations shown for Asp113 and Pro114 are also likely incorrect. The model was energy minimized with Ala117 joined to Lys1, in the same conformation as the corresponding region of PCI-1 (Pro33I-Lys34I). This was done so that the energy minimization results would correspond as closely as possible. There would presumably be no problem accomodating the additional residues at the carboxy-terminus. If one exlcudes hydrogen bonds involving side chains, and those hydrogen bonds involving residues in the area of the deletion, the two domains share a similar backbone hydrogen bonding network.

the interactions with the peptide nitrogens of Cys91 and Thr92. It should be noted, however, that the conformations shown for side chains in this model which differ from the corresponding ones in PCI-1 are somewhat arbitrary. There is no reason, then, to assume that a different set of conformations for the three threonine residues in that region (Thr88, Thr89, and Thr92) could not better satisfy these hydrogen bonding requirements. Furthermore, solvent molecules could also fulfill this role.

Despite the success of the energy minimization in obtaining a model for the trypsin domain with an energy comparable to that of PCI-1, these results must be viewed with caution. As Novotný *et al.* (1984) have shown, energy minimization does not necessarily lend any validity to the model obtained. We have shown that it is possible for the side

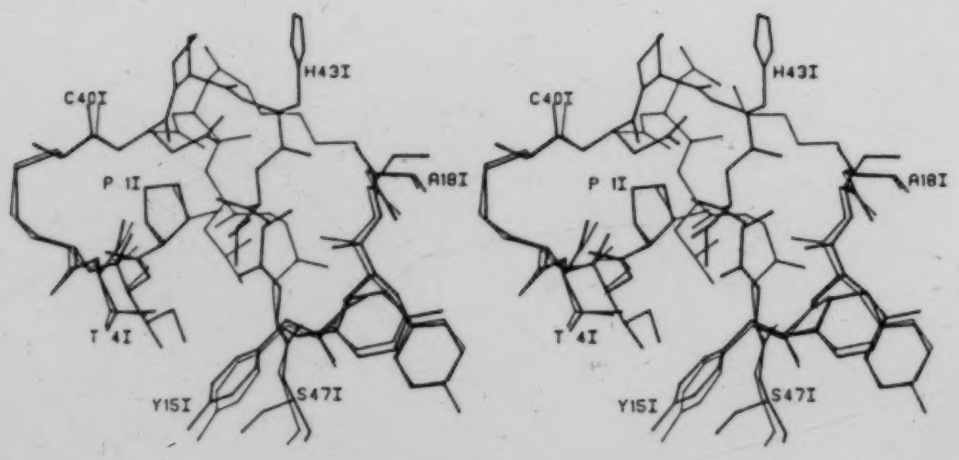


Figure III.20. Stereo view of the region surrounding the deletion between Gly8-Asn9 in the energy-minimized model of the trypsin domain of Inhibitor II. The same region of PCI-1 is drawn with thick lines, and the labels correspond to PCI-1.

chains of the residues in the trypsin domain to be accomodated in a structure similar to PCI-1. The high sequence identity (52%) between the two domains (Table 12), the conserved disulphide bridges, and the requirement for a reactive site loop do lend some credence to this model, at least in terms of its general structure.

The next step is the proposal of a reasonable association between the two domains. The positioning of the two domains relative to each other involves many variable parameters, and thus any proposals about the complete structure of Inhibitor II are speculative. Two constraints which do limit the number of possibilities are that Arg17 must be near Pro25 (ProlI of PCI-1) and likewise, Arg75 (Arg51I of PCI-1) must be near Pro84. Given these restrictions, one possible model for Inhibitor II (Figure III.21) extends the three-stranded β-sheet of PCI-1 into a six-stranded sheet with a pseudo-two fold axis roughly perpendicular to the centre of the sheet, between the two domains. Examples of similar modes of association between domains include concanavalin A (Reeke *et al.*, 1975) and HLA-A2 antigen (Bjorkman *et al.*, 1987). Since the loops which join the two domains are related by the symmetry axis, they could potentially be processed by the same

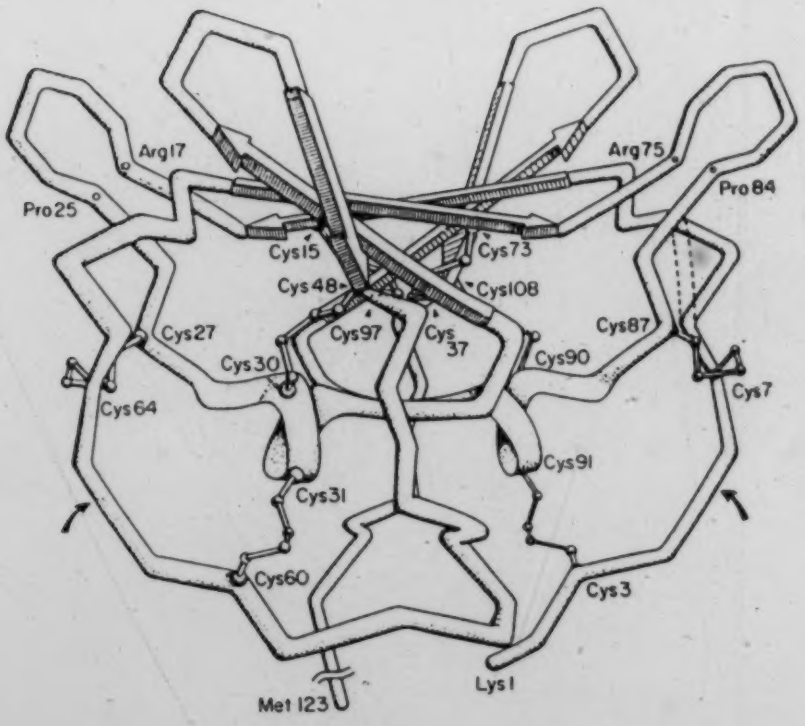


Figure III.21. A schematic diagram of the proposed model of Inhibitor II. The domain in the foreground corresponds to PCI-1, while the domain in the background contains the amino- and carboxy-termini, as well as the trypsin-specific reactive site loop. The amount of propellar twist between the two domains conveyed by the diagram is arbitrary. The reactive site loops are indicated by the curved solid arrows. Arg17 and Arg75 are the possible processing sites for the cleavages necessary to form PCI-1. Pro25 would become Pro11 of PCI-1, and Arg75 would become Arg51I. The conformations shown in the diagram for these loops are purely arbitrary, and are shown only as a visual aid. The deletion between Gly8 and Asn9 in the trypsin domain of the proline corresponding to Pro66 in the chymotrypsin domian is shown with dashed lines. The conformation of the segment of chain from Pro115 to Met123 cannot be accurately predicted, and so this region has not been shown.

mechanism. The two arginine residues, Arg17 and Arg75 are candidates for trypsin-like cleavage. One should note, however, that any serine-proteinase which might accomplish this cleavage could potentially be inhibited by the trypsin-specific reactive site loop.

The positioning of the two domains in this model is subject to several degrees of freedom. Two of the most important variables involve the interdomain angle (Figure III.22) and the propellar twist of the sheets. The interdomain angle could, in fact, be variable in the solution form of the free inhibitor. Binding of two enzymes at the two reactive site loops would likely limit this flexibility, since a small interdomain angle might result in clashes between the enzyme and the other domain or the other bound enzyme. The amount of twisting at the interface between the two domains could be very small, as in the case of HLA-A2, or it could be as large as that found in the β-sheet of PCI-1.

An alternate model for the structure of Inhibitor II involves the association of the surfaces of the β-sheets in each domain. This would give a β-barrel type structure. Difficulties arose, however, since suitable hydrogen bonds could not be found for the barrel

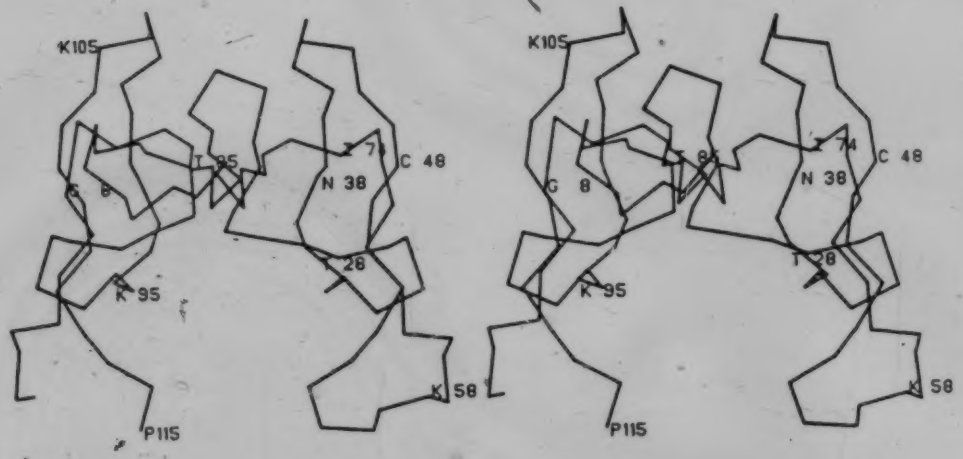


Figure III.22. C^{α} plot of Inhibitor II which is rotated ~90° about the pseudo two-fold axis relative to Figure III.21. This view demonstrates wide range of values that the interdomain angle could assume. The loops connecting the two domains are shown in arbitrary conformations, and are not meant to represent any serious attempt to model these sections. As in Figure III.21, the amount of twist in the sheets between the two domains is also arbitrary.

structure, and so the model was abandoned.

If the proposed model for the structure of Inhibitor II is correct, this protein could not have originated as a single-domain inhibitor, which, through gene duplication, became a two-domain inhibitor, since the folding would appear to be across the repeat. If gene duplication was the cause of the repeated amino acids in Inhibitor II, this process was an essential step, without which the domains could not have assumed this postulated form.

IV. Mechanism of Inhibition

A standard mechanism of inhibition for proteinase inhibitors has been discussed at length (Finkenstadt and Laskowski, 1967; Laskowski and Kato, 1980; Fujinaga *et al*, 1982; Ardelt and Laskowski, 1985; Read and James, 1986). The kinetic schemes proposed are all based upon the following equilibria:

$$E + I \rightleftarrows C \rightleftarrows E + I^{\bullet}$$

where I is the uncleaved inhibitor, $I^{\bullet}$ is the cleaved inhibitor, and C is some stable complex. Work by Ardelt and Laskowski (1985) and crystallographic evidence show that in the conditions employed in these studies the stable complex is E·I, where the intact inhibitor is tightly bound to the enzyme. The above kinetic scheme can be rewritten as

$$E + I \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} E{\cdot}I \underset{k_{on}^{*}}{\overset{k_{off}^{*}}{\rightleftarrows}} E + I^{\bullet}$$

as suggested by Read *et al* (1983). It has been argued that in this scheme, the ratio k_{off}/k_{off}^{*} is a measure of the relative heights of the activation energy barriers on the virgin and modified inhibitor sides of the reaction (Read and James, 1986).

It is useful at this point to distinguish between two mechanistic classes of inhibitors: kinetic inhibitors and thermodynamic inhibitors. A kinetic inhibitor is one that slows down hydrolysis by some means, in spite of the fact that the cleavage reaction is thermodynamically favoured. In this case, the reaction $I \rightarrow I^{\bullet}$ is spontaneous; the system $E + I^{\bullet}$ has the lowest free energy of any point along the reaction coordinate. Inhibition is achieved by raising the activation barriers at one or more stages, such that a bound form of the enzyme exists for a much longer period of time than would be observed with a regular substrate. An example of this type of inhibitor may be CI-1 (Svendsen *et al.*, 1982) which binds tightly, but is slowly cleaved and released.

A thermodynamic inhibitor is one which creates an inhibitor:enzyme complex which has the lowest free energy of any point on the reaction profile. Thus, in the presence of the enzyme, the reaction $I \rightarrow I^*$ is no longer spontaneous. Rather, the reactions $E + I \rightarrow E{\cdot}I$ and $E + I^* \rightarrow E{\cdot}I$ are thermodynamically favoured. The free energy change associated with binding of either I or I^* to an enzyme would depend upon the various interactions between the two molecules, such as hydrophobic surface area contact, and longer range electrostatic forces.

As previously noted, the reactive site loops are conformationaly constrained, either covalently by disulphide bridges (eg., PCI-1) or by other forces (eg., CI-2). (Laskowski and Kato, 1980; Fujinaga *et al*, 1982; Read *et al*, 1983; Read and James, 1986). These constraints limit the number of degrees of freedom allowed to the loop, thus decreasing its entropy relative to a peptide substrate of similar sequence. In this case, more of the enthalpy of binding is realized in the overall free energy change for the binding reaction (Blow, 1974; Read *et al*, 1983). Furthermore, it appears from work by Ardelt and Laskowski (1985), that mixing of serine proteinases with modified inhibitors (cleaved at scissile bond) produces inhibitor:enzyme complexes which have the scissile bond intact. These observations strongly suggest that the serine proteinase inhibitors are thermodynamic inhibitors, rather than kinetic inhibitors. Thus, the E·I complex is the most thermodyamically favourable state along the entire reaction coordinate.

Initially, this conclusion would appear to dispose of the need to create barriers to hydrolysis on the part of the inhibitor, as has been previously discussed. One is led, however, to question the function of the apparently conserved hydrogen bond between an acceptor and the NH of the P'_1 residue. The function of this interaction, as well as the network of hydrogen bonds that surround it in the known inhibitor crystal structures, may be to stabilize the intact inhibitor relative to the cleaved form. The enthalpic and entropic gains realized by peptide hydrolysis (Tinoco *et al*, 1978) would create a significant free enery drop in going from the uncleaved inhibitor to the cleaved form. The E·I complex would then

have to be lower in energy than the cleaved form, which may be difficult. There is, however, evidence to indicate that the equilibrium constant for the hydrolysis of the scissile peptide bond of the inhibitors is close to unity (Laskowski and Kato, 1980). This suggests that the function of the elaborate network of hydrogen bonds in the reactive site loop is to limit the entropic gain which could be realized by hydrolysis of the scissile bond. The enthalpic gain may be offest by disruption of a few of the hydrogen bonds which existed in the native inhibitor. If the free energy difference between the native and cleaved inhibitors is minimized, then association of the enzyme with either form could readily create a complex with the lowest free energy of any point on the reaction coordinate. Information to support or refute this hypothesis should be provided by the crystal structure of a free cleaved inhibitor where the extent of hydrogen bonding in the reactive site loop could be assessed. Furthermore, the structure of the complex of a cleaved inhibitor and an anhydro-enzyme (Ser195 O^{γ} removed) would provide valuable information about how the enzyme accomodates the geometric changes associated with peptide bond cleavage.

V. Conclusion

The elucidation of the structure of the PCI-1-SGPB complex has added further support to the idea that protein inhibitors of serine proteinases all interact with their cognate enzymes in similar ways, despite large differences in the structures of the inhibitors and of the enzymes. Furthermore, it appears that the problem of forming a suitable reactive site loop on the inhibitor can be solved by several structural frameworks. The combination of these two factors makes formulation of a hypothesis for the mechanism of action of these inhibitors challenging, since the same function seems to be fulfilled in a variety of ways. The mechanism of action presented in this dissertation has attempted to view the problem in a more general way than has been attempted in the past, though it is definitely based upon these earlier considerations. Furthermore, the hypothesis presented has accounted for the presence of the enzyme, and for the thermodynamics of the system as whole. This is in keeping with the issues raised by Read and James (1986), where mechanisms of inhibition which concentrated solely on the structure of the inhibitor were found to be inadequate. Nevertheless, the hypothesis presented here lacks detail, in terms of the nature of the interactions which might be lost or gained upon hydrolysis of the scissile bond. Further analysis is required, and a more adequate method of assessing the energy of interaction for different complexes is needed. Towards this end, investigations into the change in solvent accessible surface areas of various inhibitors have been initiated. It is hoped that these changes will provide more information on the energy of binding, which may then be correlated to inhibition constants.

Bibleography

Albertsson, J., Grenthe, I., & Herbertsson, H. (1973) *Acta Cryst. B29*, 1855-1860.

Ardelt, W., & Laskowski, M., Jr. (1985) *Biochemistry 24*, 5313-5320.

Baker, E. N., & Hubbard, R. E. (1984) *Prog. Biophys. molec. Biol. 44*, 97-179.

Barry, C. D., Molnar, C. E., & Rosenberger, F. U. (1976) *Technical Memo No. 229*, Computer Systems Lab, Washington University, St. Louis, MO.

Berman, H. M., Carrel, H. L., & Glusker, J. P. (1973) *Acta Cryst. B29*, 1163-1165.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L., & Wiley, D. C. (1987) *Nature 329*, 506-512.

Blow, D. M. (1974) *Bayer-Symp. 5*, 677-678.

Blundell, T. L., & Johnson, L. N. (1976) *Protein Crystallography*, New York: Academic Press.

Bode, W., Epp, O., Huber, R., Laskowski, M., Jr., & Ardelt, W. (1985) *Eur. J. Biochem. 147*, 387-395.

Bode, W., Papamokos, E., Musil, D., Seemueller, U., & Frits, H. (1986) *EMBO J. 5*, 813-818.

Bode, W., Papamokos, E., & Musil, D. (1987) *Eur. J. Biochem. 166*, 673-692.

Bolognesi, M., Gatti, G., Menegatti, E., Guarni, M., Marquart, M., Papamokos, E., & Huber, R. (1982) *J. Mol. Biol. 162*, 839-868.

Bowles, D. J., Marcus, S. E., Pappin, D. J. C., Findlay, J. B. C., Eliopoulos, E., Maycox, P. R., Burgess, J. (1986) *J. Cell Biol. 102*, 1284-1297.

Broadway, R. M., Duffey, S. S., Pearce, G., & Ryan, C. A. (1986a) *Entomol. Exp. Appl. 41*, 33-38.

Broadway, R. M., & Duffey, S. S. (1986b) *J. Insect Physiol. 32*, 827-833.

Bryant, J., Green, T. R., Gurusaddaiah, T., & Ryan, C. A. (1976) *Biochemistry 15*, 3418-3424.

Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry, Part II*, W. H. Freeman and Company, New York, 811-819.

Chambers, J. L., & Stroud, R. M. (1979) *Acta Cryst. B35*, 1861-1874.

Crowther, R. A. (1972) in *The Molecular Replacement Method* (Rossmann, M. G., Ed.) International Science Review 13, pp 173-178, Gordon & Breach, New York.

Cruickshank, D. W. J. (1947) *Acta Cryst. 2*, 65-82.

Cruickshank, D. W. J. (1967) in *International Tables for X-ray Crystallography* (Kasper, J. S., & Lonsdale, K., Eds.) Vol. 2, pp 318-340, Kynoch Press, Birmingham, England.

Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G., & Thiessen, W. E. (1975) *Nature (London) 257*, 758-763.

Einspahr, H., & Bugg, C. E. (1980) *Acta Cryst. B36*, 264-271.

Einspahr, H., & Bugg, C. E. (1981) *Acta Cryst. B37*, 1044-1052.

Einspahr, H., & Bugg, C. E. (1984) in *Calcium and its Role in Biology (Metal Ions in Biological Systems, vol.)*, (Sigel, H., ed), Marcel Dekker, New York, 51-97.

Finkenstadt W. R., Laskowski, M., Jr. (1967) *J. Biol. Chem. 242*, 771-773.

Fujinaga, M., Read, R. J., Sielecki, A. R., Ardelt, W., Laskowski, M., Jr., & James, M. N. G. (1982) *Proc. Nat. Acad. Sci. U. S. A. 79*, 4868-4872.

Fujinaga, M., Delbaere, L. T. J., Brayer, G. D., James, M. N. G. (1985) *J. Mol. Biol. 183*, 479-502.

Fujinaga, M., Read, R. J. (1987) *J. Appl. Crystallogr.20*, 517-521.

Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M., Jr., & James, M. N. G. (1987) *J. Mol. Biol. 195*, 397-418.

Graham, J. S., Pearce, G., Merryweather, J., Titani, K., Ericsson, L. H., & Ryan, C. A. (1985) *J. Biol. Chem. 260*, 6561-6564.

Hajdu, J., Machin, P. A., Campbell, J. W., Greenhough, T. J., Clifton, I. J., Zurek, S., Gover, S., Johnson, L. N., Elder, M. (1987) *Nature 329*, 178-181.

Hass, G. M., Venkatakrishnan, R., & Ryan, C. A. (1976) *Proc. Nat. Acad. Sci. U. S. A. 73*, 1941-1944.

Hass, G. M., Hermodson, M. A., Ryan, C. A., & Gentry, L. (1982) *Biochemistry 21*, 752-756.

Henderson, R. (1970) *J. Mol. Biol. 54*, 341-354.

Hendrickson, W. A. (1976) *J. Mol. Biol. 106*, 889-893.

Hendrickson, W. A., & Konnert, J. H. (1980) in *Biomolecular Structure, Function, Conformation, and Evolution* (Srinivasan, R., Ed.) Vol I, pp 43-57, Pergamon Press, Oxford.

Herzberg, O., & James, M. N. G. (1985) *Biochemistry 24*, 5298-5302.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., & Steigmann, W. (1974) *J. Mol. Biol. 89*, 73-101.

Huber, R., & Bode, W. (1978) *Acc. Chem. Res. 11*, 114-122.

Jack, A. T., & Levitt, M. (1978) *Acta Cryst. A34*, 931-935.

James, M. N. G., Brayer, G. D., Delbaere, L. T. J., Sielecki, A. R., & Gertler, A. (1980a) *J. Mol. Biol. 139*, 423-438.

James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J., & Bauer, C.-A. (1980b) *J. Mol. Biol. 144*, 43-88.

James, M. N. G., & Sielecki, A. R. (1983) *J. Mol. Biol 163*, 299-361.

Jurášek, L., Carpenter, M. R., Smillie, L. B., Gertler, A., Levy, S., & Ericsson, L. H. (1974) *Biochem. Biophys. Res. Commun. 61*, 1095-1100.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577-2637.

Kraut, J. (1977) *Ann. Rev. Biochem. 46*, 331-358.

Landon, M. (1977) *Methods Enzymol. 47*, 145-149.

Laskowski, M., Jr. & Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593-626.

Laskowski, M., Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Kohr, W. J., Park, S. J., Parks, K., Schatzley, B. L., Schoenberger, O. L., Tashiro, M., Vichot, G., Whatley, H. E., Wieczorek, A., Wieczorek, M. (1987) *Biochemistry 26*,202-221.

Mariezcurfena, R. A., & Rasmussen, S. E. (1973) *Acta Cryst. B29*, 1035-1040.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta. Cryst. B39*, 480-490.

Matsuura, Y., Kusunoki, M., Harada, W., Kakudo, M. (1984) *J. Biochem. (Tokyo) 95*, 697.

McPhalen, C. A., Schnebli, H. P., James, M. N. G. (1985a) *FEBS Lett. 188*, 55-58.

McPhalen, C. A., Svendsen, I., Jonassen, I., & James, M. N. G. (1985b) *Proc. Natl. Acad. Sci. U. S. A. 82*, 7242-7246.

McPhalen, C. A., (1986) Ph. D. Thesis, University of Alberta.

McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261-269.

McPhalen, C. A., & James, M. N. G. (1988) *Biochemistry*, in press.

Moult, J., Sussman, F., James, M. N. G. (1985) *J. Mol. Biol. 182*, 555-566.

North, A. C. T., Phillips, D. C., & Mathews, F. S. (1968) *Acta Cryst. A24*, 351-359.

Novotný, J., Bruccoleri, R., & Karplus, M. (1984) *J. Mol. Biol. 177*, 787-818.

Pähler, A., Banerjee, A., Dattagupta, J. K., Fujiwara, T., Lindner, K., Pal, G. P., Suck, D., Weber, G., & Saenger, W. (1984) *EMBO Journal 3*, 1311-1314.

Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I., & Laskowski, M., Jr. (1982) *J. Mol. Biol. 158*, 515-537.

Pearce, G., Sy, L., Russell, C., Ryan, C. A., & Hass, G. M. (1982) *Arch. Biochem. Biophys. 213*, 456-462.

Piszkiewicz, D., Landon, M., & Smith, E. L. (1970) *Biochem. Biophys. Res. Commun. 40*, 1173-1178.

Quast, U., Engel, J., Steffen, E., Tschesche, H., & Kupfer, S. (1978) *Eur. J. Biochem. 86*, 353-360.

Quiocho, F. A., Sack, J. S., & Vyas, N. K. (1987) *Nature 329*, 561-564.

Ramachandran, G. N., & Mitra, A. K. (1976) *J. Mol. Biol. 107*, 85-92.

Ramakrishnan, C., & Ramachandran, G. N. (1965) *Biophys. J. 5*, 909-933.

Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420-4433.

Read, R. J. (1986a) *Acta. Cryst. A42*, 140-149.

Read, R. J. (1986b), Ph. D. Thesis, University of Alberta.

Read, R. J., & James, M. N. G. (1986) in *Proteinase Inhibitors, chap 7* (Barrett and Salvesen, Eds.), Elsevier Science Publishers BV (Biomedical Division), 301-336.

Read, R. J., & James, M. N. G. (1988) *J. Mol. Biol. 199*, in press.

Reeke, G. N., Jr., Becker, J. W., & Edelman, G. M. (1975) *J. Biol. Chem. 250*, 1525-1547.

Rees, D. C., Lewis, M., & Lipscomb, W. N. (1983) *J. Mol. Biol. 168*, 367-387.

Richardson, M. (1979) *FEBS Lett. 104*, 322.

Richardson, J. S., Getzoff, E. D., Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. USA 75*, 2574-2578.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., & Wilcox, P. E. (1972) *Biochemistry 11*, 2439-2449.

Rossmann, M. G. (ed) (1972) *The Molecular Replacement Method*, International Science Review 13, Gordon & Breach, New York.

Ryan, C. A., Bishop, P. D., Walker-Simmons, M., Brown, W. E., & Graham, J. S. (1985) in *Cellular and Molecular Biology of Plant Stress* (J. King & Tsune Kosuge, Eds.), pp 319-334, Allan R. Liss Inc., New York.

Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157-162.

Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G.

D., & James, M. N. G. (1979) *J. Mol. Biol. 134*, 781-804.

Sielecki, A. R., James, M. N. G., & Broughton, C. G. (1982) in *Crystallographic Computing, Proceedings of the International Summer School* (Sayre, D., Ed.) pp 409-419, Carelton University, Ottawa, Oxford University Press, Oxford.

Steitz, T. A., & Shulman, R. G. (1982) *Ann. Rev. Biophys. Bioeng. 11*, 419-444.

Stout, G. H., & Jensen, L. H. (1968) *X-ray Structure Determination*, Macmillan, New York.

Streitweiser, A., Jr., & Heathcock, C. H., (1981) *Introduction to Organic Chemistry (2nd Edition)*, Macmillan, New York, 540-542.

Svendsen, I., Boisen, S., & Hejgaard, J. (1982) *Carlsberg Res. Commun. 47*, 45-53.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., & Blow, D. M. (1974) *Biochemistry 13*, 4212-4228.

Takio, K., Blumenthal, D. K., Walsh K. A., Titani, K., & Krebs, E. G. (1986) *Biochemistry 25*, 8049-8057.

Thiessen, W. E., & Levy, H. A. (1973) *J. Appl. Crystallogr. 6*, 309.

Thornburg, R. W., An, G., Cleveland, T. E., Johnson, R., Ryan, C. A. (1987) *Proc. Natl. Acad. Sci. U. S. A. 84*, 744-748.

Tinoco, I., Jr., Sauer, K., Wang, J. C. (1978) *Physical Chemistry, Principals and Applications in Biological Sciences*, pp. 79-81, Prentice-Hall, Englewood Cliffs, N. J.

van Gunsteren, W. F. & Berendsen, H. J. C. (1984). Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

Weber, E., Papamokos, E., Bode, W., Huber, R., Kato, I., & Laskowski, M., Jr. (1981) *J. Mol. Biol. 149*, 109-123.

Wilmot, C. M., & Thornton, J. M. (1988) *ICSU Short Reports, Advances in Gene Technology: Protein Engineering and Production 8*, 20.